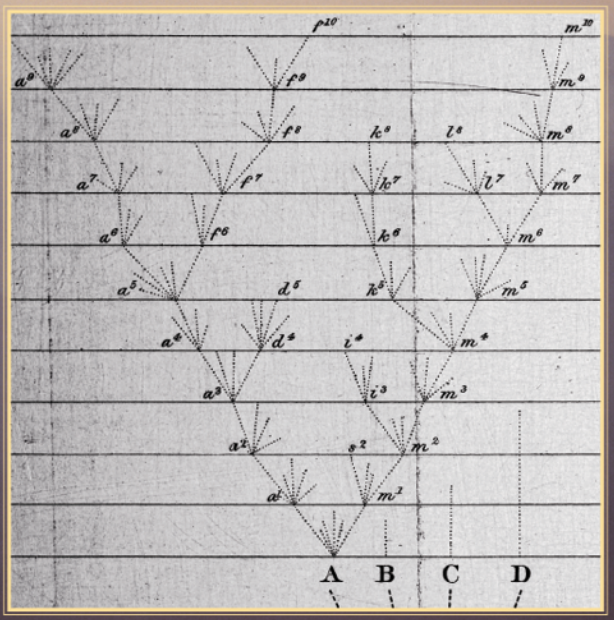
Handbook of
THE PHILOSOPHY OF SCIENCE
General Editors: DOV M. GABBAY, PAUL THAGARD, AND JOHN WOODS

PHILOSOPHY *of* BIOLOGY



Edited by Mohan Matthen
and Christopher Stephens

N·H

Philosophy of Biology

Handbook of the Philosophy of Science

General Editors

Dov M. Gabbay
Paul Thagard
John Woods

Cover figure taken from: Darwin's Origin of Species, 1859

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
North-Holland is an imprint of Elsevier

Philosophy of Biology

Edited by

Mohan Matthen
University of Toronto, Canada

and

Christopher Stephens
University of British Columbia, Canada

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
North-Holland is an imprint of Elsevier

North-Holland is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2007

Copyright © 2007 Elsevier B.V. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at http://elsevier.com/locate/permissions, and selecting *Obtaining permission to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN-13: 978-0-444-51543-8
ISBN-10: 0-444-51543-7

For information on all North-Holland publications
visit our website at books.elsevier.com

Printed and bound in The Netherlands

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

GENERAL PREFACE

Dov Gabbay, Paul Thagard, and John Woods

Whenever science operates at the cutting edge of what is known, it invariably runs into philosophical issues about the nature of knowledge and reality. Scientific controversies raise such questions as the relation of theory and experiment, the nature of explanation, and the extent to which science can approximate to the truth. Within particular sciences, special concerns arise about what exists and how it can be known, for example in physics about the nature of space and time, and in psychology about the nature of consciousness. Hence the philosophy of science is an essential part of the scientific investigation of the world.

In recent decades, philosophy of science has become an increasingly central part of philosophy in general. Although there are still philosophers who think that theories of knowledge and reality can be developed by pure reflection, much current philosophical work finds it necessary and valuable to take into account relevant scientific findings. For example, the philosophy of mind is now closely tied to empirical psychology, and political theory often intersects with economics. Thus philosophy of science provides a valuable bridge between philosophical and scientific inquiry.

More and more, the philosophy of science concerns itself not just with general issues about the nature and validity of science, but especially with particular issues that arise in specific sciences. Accordingly, we have organized this Handbook into many volumes reflecting the full range of current research in the philosophy of science. We invited volume editors who are fully involved in the specific sciences, and are delighted that they have solicited contributions by scientifically-informed philosophers and (in a few cases) philosophically-informed scientists. The result is the most comprehensive review ever provided of the philosophy of science.

Here are the volumes in the Handbook:

Philosophy of Science: Focal Issues, edited by Theo Kuipers.

Philosophy of Physics, edited by John Earman and Jeremy Butterfield.

Philosophy of Biology, edited by Mohan Matthen and Christopher Stephens.

Philosophy of Mathematics, edited by Andrew Irvine.

Philosophy of Logic, edited by Dale Jacquette.

Philosophy of Chemistry and Pharmacology, edited by Andrea Woody and Robin Hendry.

Philosophy of Statistics, edited by Prasanta S. Bandyopadhyay and Malcolm Forster.

Philosophy of Information, edited by Pieter Adriaans and Johan van Benthem.

Philosophy of Technological Sciences, edited by Anthonie Meijers.

Philosophy of Complex Systems, edited by Cliff Hooker and John Collier.

Philosophy of Earth Systems Science, edited by Bryson Brown and Kent Peacock.

Philosophy of Psychology and Cognitive Science, edited by Paul Thagard.

Philosophy of Economics, edited by Uskali Mäki.

Philosophy of Linguistics, edited by Martin Stokhof and Jeroen Groenendijk.

Philosophy of Anthropology and Sociology, edited by Stephen Turner and Mark Risjord.

Philosophy of Medicine, edited by Fred Gifford.

Details about the contents and publishing schedule of the volumes can be found at http://www.johnwoods.ca/HPS/.

As general editors, we are extremely grateful to the volume editors for arranging such a distinguished array of contributors and for managing their contributions. Production of these volumes has been a huge enterprise, and our warmest thanks go to Jane Spurr and Carol Woods for putting them together. Thanks also to Andy Deelen and Arjen Sevenster at Elsevier for their support and direction.

CONTENTS

General Preface v
Dov Gabbay, Paul Thagard, and John Woods

Preface xi
Mohan Matthen and Christopher Stephens

List of Contributors xv

I. Biography

Charles Darwin 1
Michael Ruse

Sir Ronald Aylmer Fisher 37
Robert A. Skipper, Jr.

Haldane and the Emergence of Modern Evolutionary Theory 49
Sahotra Sarkar

Sewall Wright 87
James F. Crow

Motoo Kimura 101
James F. Crow

II. Evolution

Natural Selection 111
Christopher Stephens

Neutralism 129
Anya Plutynski

Levels of Selection 141
Robert A. Wilson

What is Evolvability? 163
Kim Sterelny

Development: Three Grades of Ontogenetic Involvement 179
D. M. Walsh

Evolution and Normativity 201
Michael Bradie

Evolutionary Ethics 219
Catherine Wilson

III. Genetics

Genetic Analysis 249
Raphael Falk

The Development of Population Genetics 309
Margaret Morrison

Maximisation Principles in Evolutionary Biology 335
A. W. F. Edwards

Reductionism in Biology 349
Alex Rosenberg

Traits, Genes, and Coding 369
Michael Wheeler

IV. Taxonomy

Species, Taxonomy, and Systematics 403
Marc Ereshefsky

Homology and Homoplasy 429
Brian K. Hall

Biological Conceptions of Race 455
Robin O. Andreasen

V. Special Topics

Formalisations of Evolutionary Biology 485
Paul Thompson

Functions 525
Tim Lewens

Biological Approaches to Mental Representation 549
Karen Neander

Innateness 567
André Ariew

Artificial Life 585
Mark A. Bedau

Index 605

This page intentionally left blank

PREFACE

Philosophy of Biology has taken flight in the last quarter of the twentieth century. In very large part, of course, this is because biology itself has made huge strides in this period of time. Not only has it taken large strides, but it has done so in a way that it has made it a player in many of the largest public issues of the day: DNA, cloning, evolution, evolutionary psychology. It is not an exaggeration either to say that Darwin, cloning, and DNA now occupy as much space in academic and intellectual conversations as Shakespeare or the Bible. This has had a dramatic effect on philosophers. What kind of process is evolution? Are morals and meaning reducible to biology? What is the significance of genetic inheritance? Of the molecular substrate of life? Questions like these bubble up from common discourse, from popular culture, from discussions in coffee shops. And philosophers have been willing and well-equipped to take them up.

As recently as 1975, philosophy of biology was not generally recognized as a subfield, though texts by Michael Ruse and David Hull had appeared. Many of those whom we recognize today as the most prominent philosophers of biology were engaged in other things: Elliott Sober was writing about simplicity, Daniel C. Dennett about Rylean theories of mind, Philip Kitcher about mathematics. Virtually no University Department thought: "What we really need right now is a philosopher of biology." Philosophical discussions of biological issues were only just beginning to go beyond the rather mechanical business of localizing debates emanating from general philosophy of science. The pioneers were in place, of course. Logical empiricists like Carl Hempel and Ernest Nagel had discussed teleology. David Hull and Nicolas Jardine had done some startlingly original work on biological systematics. Arthur Burks, the visionary from Michigan, was provoking interest in genetic algorithms. Marjorie Grene and Michael Ruse blazed trails in discussions of evolution (not to mention Karl Popper in his peculiarly opinionated way). There was desultory interest in philosophical issues surrounding the reduction of Mendelian genetics to molecular biology, though here the general assumption was that the issues were really no different from those in the reduction of thermodynamics to statistical mechanics. Some of the philosophically interesting work in this period took place in historical discussions — in the work of such Aristotelian scholars as David Balme and Geoffrey Lloyd on taxonomy, for example, and in T. S. Hall's *Ideas of Life and Matter*. And then, of course, there were those outside philosophy, writing in ways that were recognizably philosophical: Richard Dawkins, E. O. Wilson, Richard Lewontin, Stephen Gould, Michael Ghiselin, and

Donald Campbell are scholars from whom we can all still learn. Further afield, there were the vitalists, the mysterians, the Bergsonians. But there was nothing like a discipline, no agreed upon central issues or standards or reading lists. And no paths were recognized as dangerous to traverse. That was how things were in the seventies: a thousand flowers bloomed, but it was impossible, or at the very least dangerous, to try it on in graduate school.

Things have changed. There is now a clearly defined dialectic and many more scholars who contribute to it. Nonetheless, because the field is so young, it is difficult to define problem areas and dialectics that will be valid twenty years from now — and this is what we took our job to be as editors of this volume. Broadly speaking, there are two kinds of inquiry that we wanted to include: first, the conceptual investigation of biological problems, and second, the ways that biology has been a source of illumination for problems outside of the philosophy of biology itself. It is perhaps in the second area that the "new biology" has most significance. Physics is never going to throw much light on human nature, or the nature of cognition, or of society, and so philosophers of physics do not have to concern themselves with these things. But biology clearly intrudes into these areas of inquiry, legitimately or not, and it is the task of scientifically well-informed philosophers and philosophically minded scientists to probe, monitor, police, and regulate these studies. On the other hand, the illumination is reciprocal: the applications of biology tend to confirm or disconfirm the ambitions and pretensions of biological speculations. For these reasons, the philosophy of biology is much broader than biology itself. It is concerned with conceptual analysis, but with an eye to a broad range of issues.

Evolution occupies a great deal of this Handbook. In the part explicitly devoted to the topic, Christopher Stephens deals with broad issues of explanation and confirmation in the theory of natural selection, Anya Plutynski with neutralism (i.e., evolution in which the intensity of selection is relatively muted), Robert A. Wilson with the units of selection debate, and Kim Sterelny with the relatively new topic of evolvability. This part of the Handbook concludes with two essays, by Michael Brady and Catherine Wilson, on norms that arise out of evolution.

Evolution is also crucial to the discussion of other areas of biology - such as the nature of development, dealt with by Denis Walsh in Part I, and taxonomy, dealt with by Marc Ereshefsky in Part IV, as well as the closely related issue of race, which is the topic of Robin O. Andreason's entry in Part IV. In evolutionary theory, there has been a recrudescence, in recent years, of Darwin's essentially historical approach to explaining the origin of species — which for a while was neglected in favour of studies of adaptation and optimization (see A. W. F. Edwards' entry) in the middle part of the last century. Part IV explores this in the context of taxonomy, where in addition to the essays by Ereshefsky and Andreason, Brian Hall provides an introduction to the foundations of this historical approach in his entry on Homology and Homoplasy.

Then there is genetics. In Part III, Rafi Falk traces the emergence and development of the gene-concept in his entry "Genetic Analysis". Margaret Morrison

recounts the emergence of population genetics. (Her discussion is complemented by that of Sahotra Sarkar in his entry on Haldane.) Alexander Rosenberg and Michael Wheeler discuss molecular genetics, Rosenberg dealing with issues of reduction, and Wheeler digging into the notion of genetic coding.

Part V concerns a group of essays dealing with philosophical applications arising out of the practice of biology. Here, Paul Thompson deals with formal systems that are meant to capture the content of biological theory; Tim Lewens writes on functions, Karen Neander on mental representation, André Ariew on innateness, and Mark Bedau on artificial life.

The reader will not fail to notice the orientation of Part I, which is devoted to intellectual biography. For as we were planning this section, we found that our attention came willy-nilly to be focussed entirely on one interpretation of how evolution is important. Thus, we included not only Charles Darwin in this section (the entry is by Michael Ruse), but also the four great figures of population genetics, R. A. Fisher (Robert Skipper), Sewall Wright (Jim Crow), J. B. S. Haldane (Sahotra Sarkar), and Motoo Kimura (Jim Crow again). In so doing, we left out not only such fascinating figures as Gregor Mendel (but see Falk), James Watson, and Francis Crick, but also competing figures in the study of evolution such as Ernst Mayr (whose bitter disagreement with the heroes of population genetics is recorded in the last ten pages of Sahotra Sarkar's essay on Haldane). In part, this has to do with the theoretical character of this branch of evolution — population genetics deals with abstractions, and as such it attracts the attention of philosophers more than natural history does. More relevantly, it reflects actuality: what philosophers of biology want to write about. No doubt, this will change; for in time, questions about genetics may come to occupy more of the foreground. When that happens, the intellectual biographies of figures such as Thomas Hunt Morgan and Herman J. Müller may occupy the attention of more scholars within philosophy. For now, we can only hope that the issue-oriented histories presented by Falk and Morrison will serve to provoke more philosophical discussion of the history of genetics.

We canvassed a large number of people in the field, and we learned a great deal about what occupied their attention. We learned, for example, that creationism just does not occupy the attention of philosophers of biology the way it did even a few years ago. While not everything that needs to be said about this subject has been said, many philosophers now view it as more of social and political than of philosophical interest. As it turned out, the Handbook is pre-eminently focussed on historical and conceptual issues surrounding evolutionary theory. This is so for two reasons. First, evolution is arguably at the conceptual centre of contemporary biology, with a number of particularly interesting philosophical dimensions. Second, most of the scholars of the last few decades have in fact focussed on the conceptual, methodological and historical issues in evolutionary theory.

In this sense our anthology reflects the bias of the interests of these scholars. We would hope and expect that in another few decades, the field will continue to grow and look considerably different than it does now, and that if we were to commission scholars for such a Handbook in 2050, we could look back and see that progress had been made on a number of other issues, and that new topics and approaches to old problems were revealed, just as looking back now we see the progress that has been made in approaches to evolution even in the last thirty years.

We should like to thank John Woods for encouragement and advice, and Jane Spurr for her superb (and very fast) job of editing.

Mohan Matthen
Christopher Stephens

October 2006

CONTRIBUTORS

André Ariew
Department of Philosophy, 438 General Classroom Bldg., University of Missouri-Columbia, Columbia, MO 65211-4160, USA.
ariewa@missouri.edu

Robin O. Andreasen
University of Delaware, Departments of Philosophy and Linguistics, Newark, DE 19716, USA.
robina@udel.edu

Mark A. Bedau
Reed College, 3203 SE Woodstock Blvd., Portland OR 97202-8199, USA.
mab@reed.edu

Michael Bradie
Department of Philosophy, Bowling Green State University, Bowling Green, OH 43403, USA.
mbradie@bgsu.edu

James F. Crow
Genetics Laboratory, University of Wisconsin, Madison, Wisconsin 53706, USA.
jfcrow@wisc.edu

A. W. F. Edwards
Gonville and Caius College, University of Cambridge, Cambridge, CB2 1TA
awfe@cam.ac.uk

Marc Ereshefsky
Department of Philosophy, University of Calgary, Calgary, Alberta T2N 1N4, Canada.
ereshefs@ucalgary.ca

Raphael Falk
Department of Genetics, The Hebrew University, Edmond Safra Campus - Givat Ram Jerusalem 91904, Israel.
rfalk@cc.huji.ac.il

Brian Hall
Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1.
bkh@dal.ca

Tim Lewens
Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge, CB2 3RH.
tml1000@cam.ac.uk

Mohan Matthen
Department of Philosophy, University of Toronto at Mississauga, 3359 Mississauga Rd N., Mississauga, ON L5L 1C6, Canada.
mohan.matthen@ubc.ca

Margaret Morrison
Trinity College, University of Toronto, Toronto, Ontario, M5S 1H8, Canada.
mmorris@chass.utoronto.ca

Karen Neander
Department of Philosophy, 201 West Duke Building, Duke University, Durham, North Carolina 27708, USA.
kneander@duke.edu

Anya Plutynski
Department of Philosophy, University of Utah, 260 S. Central Campus Drive, Orson Spencer Hall, Salt Lake City, UT 84112, USA.
plutynski@philosophy.utah.edu

Alex Rosenberg
Department of Philosophy, Center for Philosophy of Biology, 203-A West Duke Building, Duke University, Durham, NC 27708, USA.
alexrose@duke.edu.

Michael Ruse
Department of Philosophy, Florida State University, Tallahassee, FL 32306-1500, USA.
mruse@mailer.fsu.edu

Sahotra Sarkar
The University of Texas at Austin, Philosophy, 1 University Station C1200, Austin, Texas 78712, USA.
sarkar@mail.utexas.edu

Robert Skipper
Department of Philosophy University of Cincinnati P.O. Box 210374 Cincinnati, OH 45221-0374, USA.
robert.skipper@uc.edu

Christopher Stephens
Department of Philosophy, University of British Columbia, 1866 Main Mall E370, Vancouver BC Canada V6T 1Z1.
cstephen@interchange.ubc.ca

Kim Sterelny
School of History, Philosophy, Political Science, and International Relations, Victoria University of Wellington PO Box 600 Wellington 6001, New Zealand; and Philosophy Program RSSS, Australian National University, 0200, ACT, Australia.
Kim.Sterelny@vuw.ac.nz; kimbo@coombs.anu.edu.au

Paul Thompson
Institute for the History and Philosophy of Science and Technology and Department of Ecology and Evolutionary Biology, University of Toronto, 91 Charles Street West, Toronto, ON, Canada M5S 1K7.
p.thompson@utoronto.ca

Denis Walsh
Department of Philosophy and Institute for the History and Philosophy of Science and Technology, University of Toronto, 215 Huron Street, Toronto ON, Canada, M5S 1A2.
denis.walsh@utoronto.ca

Michael Wheeler
Department of Philosophy, University of Stirling, Stirling, FK9 4LA, Scotland.
m.w.wheeler@stir.ac.uk

Catherine Wilson
The Graduate Center, The City University of New York, 365 Fifth Avenue, New York, NY 10016-4309, USA.
cwilson@gc.cuny.edu

Robert A. Wilson
Department of Philosophy, 4-115 Humanities Centre, University of Alberta, Edmonton, Alberta, Canada T6G 2E5.
rob.wilson@ualberta.ca

This page intentionally left blank

Part I
Biography

This page intentionally left blank

CHARLES DARWIN

Michael Ruse

Charles Robert Darwin was born in the English Midlands county of Shropshire on February 12, 1809, the same day as Abraham Lincoln across the Atlantic. He died at his home in Kent, south of London, on April 19, 1882. The fourth child, second son, he was born into a distinguished and wealthy, upper-middle-class family. His paternal grandfather, Erasmus Darwin, was well known as a brilliant physician, friend of industrialists, and popular poet. His father, Robert Darwin, was also an excellent physician, as well as being a money man. He would arrange mortgages between landed gentry and aristocrats, with need of cash and land to mortgage, and industrialists looking for safe ways to invest their earnings. On his mother's side, the line was even more distinguished, for Darwin's maternal grandfather was Josiah Wedgwood, who had industrialized the British pottery trade. Darwin further sealed the Wedgwood connection (and supply of cash) when he married his first cousin, Emma Wedgwood.[1] It is important from the start to emphasize both the family status and wealth, to explain why Darwin never held a paid job throughout his life — he had no need to (and he was a canny investor) — and to make clear that Charles Darwin was never going to be the Christian God. He was not about to reject his background. He was a genius of the first order, but his genius would be to take what he had been given and to make of it a new picture — rather like a kaleidoscope being rotated. Darwin was not in the business of making something out of nothing.

Darwin was sent as a teenager to one of England's leading private (confusingly known as "public") schools, Shrewsbury. From there, at the age of 17 he was packed off to Edinburgh University, in the path of his father and grandfather, to train as a physician. He hated the course, and two years later transferred down to Cambridge University intended to become a Church of England parson. On graduating in 1831, however, his clerical plans were postponed (and then shelved) when through a friend of a friend he was offered a place on *HMS Beagle*, then about to embark on a trip to South America in order to chart the waters. In all, Darwin spent some five years on the *Beagle*, returning to England in the fall of 1836. One result of the voyage is that he wrote up his diary as a narrative, and this was published and became one of the most-loved travel books in Victorian

[1]The most reliable and detailed biography of Charles Darwin is the two volume work by Janet Browne [1995; 2002]. The most fun to read, although highly unreliable given the authors' Marxist leanings, is by Adrian Desmond and James Moore [1992]. Even the title is over the edge. Darwin's own autobiography is not always reliable but a good read nevertheless [Darwin, 1969]. I have written on the Darwinian Revolution in Ruse [1979, new edition 1999a]. More generally on the history of evolutionary thought, consult Ruse [19960;, 1999b;, 2003;, 2005]).

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

England [Darwin, 1845]. By the mid-1840s, "Darwin of the *Beagle*" was one of the better-known figures in the country. The warmth and sheer love of travel and of nature infected his fellow citizens, and by mid-century he was well on the way to being one of the people of whom the English were genuinely affectionate and proud. This incidentally continues to this day. Darwin's portrait is on the back of the ten pound note.

For reasons that are still not completely understood, shortly after the *Beagle* voyage Darwin fell sick with an illness that plagued him for the rest of his life. Headaches, insomnia, boils, bad breath, bowel upsets, and more. Some think it was psychological — being the author of a theory that was to be so controversial or perhaps fearing his father's disapproval (despite the fact that Darwin always claimed to love and admire his father) — and others think it was physical — perhaps Chagas Disease, picked up from an insect bite when he was traveling across the Andes. Whatever the cause, Darwin became an invalid and near recluse, retiring with his new wife to the village of Downe, where they proceeded to have a typically large Victorian family. Ten children were born, and seven lived to maturity. The most famous was George Darwin, who became a leading authority on the tides. The most loved was Annie, who died at the age of ten, leaving Darwin and his wife with a hole never to be filled.

Although isolated, Darwin carried on a tremendous correspondence, and (clearly using his illness to his own advantage) worked non-stop at projects that interested him. He was also good at networking, and encouraged young men whose interests he shared and (more importantly) whom he saw as those who would push his ideas. Notably in this group were the botanist Joseph Hooker and the morphologist Thomas Henry Huxley (the grandfather of Aldous Huxley, the novelist). Building on his fame as a travel guide, Charles Darwin increasingly became known both to professional scientists and to laypeople as one of the truly great men of his time. It was therefore absolutely no surprise whatsoever that when he died, by the overwhelming demand of the country, Charles Darwin was buried in the English Valhalla, Westminster Abbey. There he lies to this day, right next to Sir Isaac Newton.

DARWIN THE GEOLOGIST

Let us start now to look at Charles Darwin as a scientist. His own *Autobiography* is somewhat misleading on this issue. There he presents himself as a rather slow fellow, who succeeded in spite of his deficiencies. But we must not let that upper-middle-class modesty — a group where "swanking" is the worst of all possible sins — mislead us. It is clear that Darwin was no very great shakes as a scholar when he was at school. It is clear also that Darwin (unlike his son George) was never very good at mathematics. However, this said, negatively we should know that English education in those days was an uninterrupted diet of Latin and Greek, with Euclid for the truly gifted, and Darwin was not of this nature. Positively we know that even as a teenager, Charles Darwin and his older brother Erasmus

tried out various chemical experiments in a shed at the bottom of the garden. No surprise here. Chemistry is the industrial science *par excellence*, and the Darwin-Wedgwood family were doing very well out of industry. When Darwin went off to Edinburgh, he may have hated the courses in medicine — one major factor with which anyone could sympathize was the need to rise early in the morning, in the winter long before daybreak, to listen to Scots professors on pharmacology — but he started to mix with people interested in natural history, biology as we would call it today. One figure in the group was Robert Grant, one of the few evolutionists of the day.

The pattern continued down at Cambridge. Becoming a clergyman for one such as Darwin was the ideal profession. It was respectable, it did not require too much work (especially if one could afford to hire a curate to do the hard jobs), and (to speak very euphemistically) it was not a post that required an abundance of grey matter. It was the perfect job for a gentleman. Darwin was not expected to work too hard at his studies and he did not. Enough to get through respectably. However, he continued with his interests in natural history — collecting beetles, attending the lectures of the Professor of Botany, John Henslow, and talking to those few men who were trying to make Cambridge more science friendly. The offer of the *Beagle* voyage at the end of the time was no chance. Initially Darwin was intended to be basically a companion for the captain, but soon he matured into the ship's naturalist, sending massive collections back home to England.

Darwin's first major scientific forays were into geology [Herbert, 2005]. This was a much-discussed science around the 1830s. Like chemistry, it was a major factor in the Industrial Revolution. Minerals were needed, and coal also. No one wanted to spend a fortune on sinking mines in areas that were simply not going to be productive. Roads were being built, and canals were the major way in which goods could be transported. Again, no one wanted to try to bore a tunnel through a mountain that proved to be granite rather than sandstone. Railways were just around the corner. Hence interest in geology was high, and the scientific group to which Darwin had attached himself was in the thick of things.

There were two major theories about geological change. On the one hand, named by the scientific polymath William Whewell [1837], there were the *catastrophists*. Much influenced by the thinking of the French comparative anatomist, Georges Cuvier [1813], they argued that every now and then there are mighty convulsions, and after these cataclysmic events organisms are recreated miraculously, to meet the new conditions. Earth's history is directional, from hot to cold, and leads to the world that we have today. On the other hand, there were those that Whewell labeled *uniformitarians*. Their leader was Charles Lyell, and in his three-volume *Principles of Geology* [1830–1833] he argued that the same laws and effects hold throughout geological history — rain, sleet, wind, volcanoes, earthquakes, and so forth can do it all. The only requirement is masses and masses of time.

Darwin might have been expected to attach himself to the catastrophic camp. His teacher of geology was a leading catastrophist, Adam Sedgwick, professor of

Geology at Cambridge University. But probably in part because of the isolation on the *Beagle* — Darwin took with him the first volume of Lyell, and the others were sent out on the voyage — he became an ardent uniformitarian. This set the scene for two major pieces of geological work that Darwin did – one successful and one very much non-successful. Lyell wanted to deny not just catastrophes but also (for reasons to be given shortly) the directional nature of the world. How then could he account for the seeming effects of cooling? The fossil plants around Paris are tropical, clear evidence that things are not as they once were. Lyell introduced his "grand theory of climate." It is a mistake to think that temperature is a function of distance from the equator. It is rather a function of relative distributions of land and sea. Consider Britain, much warmer than it should be, thanks to the Gulf Stream. What happens is that the world acts rather like a water bed. (This is my analogy, not his!) As the rivers carry silt from the mountains down to the sea, where it is deposited, the sea bed starts to sink because of the weight, whereas the mountains start to rise because they are not as heavy as previously. Things shift around, and then natural processes affect the terrain in new ways. The consequence is that, because of Gulf Stream effects, the temperature around the globe keeps shifting. Never the same, always moving up and down, but within limits, rather like a sine curve.

Darwin bought into this completely, and his two major geological theories were in fact responses to challenges posed by Lyell in the *Principles*. First there is the question of coral reefs [Darwin, 1842]. In the tropics, one often finds circular reefs around islands, rings of coral. Sometimes, indeed, there are no islands and just the reefs. How can this be? Somewhat optimistically, Lyell suggested that they might be the tops of extinct volcanoes. The coral just forms on them. The great weakness here — one that Darwin seized on — is that it is highly improbable that so many volcanoes would have risen from the sea bed just to the point where they break the water's surface. Darwin therefore turned the problem on its head. He pointed out that coral can only live at the surface of the water. Hence he suggested that initially the coral grows around the edges of islands, and then its weight causes everything to start sinking. The islands go down and perhaps vanish beneath the sea, but the coral keeps building on what is there and so stays visible at the surface.

This was a triumph of reasoning. Darwin admitted that he had thought it all up even before he had seen a coral reef! Today it is accepted as true. Borings of coral reefs bear out Darwin's hypothesis that they go all the way down. Notice, however, to what extent it was a corollary of Lyellian geology. Precisely what Thomas Kuhn [1962] would have called normal science within the paradigm. The same is true of Darwin's [1839] second major piece of geologizing, the so-called Parallel Roads of Glen Roy. Along the sides of one of the small valleys off the Great Glen in Scotland, there are parallel tracks. What could have caused them? General opinion was that they were not human-caused — old hunting tracks — but natural, being the remains of shores of water once filling the glen. But where is the water now?

There were two hypotheses. One was that there had been a lake that was now gone. The other was that the sea had once run into the glen but no longer. After he arrived back from the voyage, Darwin dashed up to Scotland to look at the glen for himself. He was already predisposed to think that the sea-hypothesis was the right one. This was the speculation in the *Principles*, and it fit right into the theory of climate. The land in Scotland around the Great Glen had been rising, and this meant that the sea ran out of Glen Roy. Darwin expected to find remains of marine organisms on the tracks, but the failure to find such specimens did not deter him. After all, if it had been a lake, where now is the barrier that kept it in? This work resulted in Darwin's one and only one paper in the prestigious *Philosophical Transactions of the Royal Society* — Sedgwick was one of the (favourable) referees, showing that the uniformitarian/catastrophist divide was not personal — but alas it was all a massive mistake. A year or two later, the Swiss ichthyologist Louis Agassiz (he who was later to go to Harvard and become a bitter critic of Darwinism) visited Scotland. Used to glaciers and skilled in studying rocks for their evidence, he immediately announced (what is true) that the glen had held a lake, kept in place by a glacier that has since melted. Many years later Darwin agreed that he had been wrong. The length of time that it took for Darwin to admit his mistake was not simply a function of pride. Agassiz was tying in his glacier work with the thesis that there had been ice ages — which was fair enough — evidence that he took of God's intervening powers — which was not fair enough [Rudwick, 1974].

DARWIN BECOMES AN EVOLUTIONIST

At least not fair enough in the eyes of a Lyell or a Darwin. We move towards our main story, Darwin and evolutionary theory. But as we do so, let us take a moment to reflect on religion, both generally at the time and specifically in the mind of Charles Darwin. No one but crackpots on the fringe took Genesis absolutely literally. The Bible was true throughout but those early bits were to be understood in a metaphorical or allegorical sense. Although there had been speculation on the possibility of Noah's Flood, by the time that Darwin went to Cambridge that was denied. The catastrophists were certainly closer to a literal reading than the uniformitarians — some thought the six days were six long periods of time, whereas others thought there were unexplained passages of time unmentioned in Genesis — but everyone thought the earth very old. How old, catastrophists generally preferred not to say, on the sensible grounds that being specific could only lead to trouble.

As an undergraduate, it seems that Darwin was really quite conservative theologically. Indeed at first on the *Beagle* the officers used to twit him for his literal beliefs. But shortly after setting off, Darwin's views started to change. He began as a Christian theist — that is as someone who believes that God interferes in His creation (most obviously in the Incarnation) — but evolved into a deist — that is as someone who believes in God as unmoved mover. God set things going and

then lets unbroken law do the job. In other words, Darwin moved from seeing miracles as the greatest evidence of God's existence (something he would have gleaned from reading Archdeacon Paley's standard text, *Evidences of Christianity* [1794], when at Cambridge) to seeing laws as the greatest evidence of God's existence (something he would have gleaned from reading Charles Lyell's *Principles of Geology* on the voyage).

One should stress that this was no great move for Darwin, and certainly not something that would have been socially ostracizing. Lyell was openly a deist, at that time worshiping with the Unitarians (a group who, in those days, believed in God but denied the Trinity and hence the intervention of the Incarnation). Darwin's grandfather Erasmus was a deist, and the whole of the Wedgwood family were Unitarians. So really Darwin was just changing one family belief for another. (Darwin's father Robert was probably an out-and-out atheist, but for professional reasons obviously had to keep his non-belief under cover. He was certainly cynical — or sensible depending on your viewpoint — in packing Charles off to Cambridge to become a parson. But what else was a concerned father to do when faced with an undirected son?)

It is probable that deism stayed with Darwin for many years, but then late in life his beliefs drifted into the agnosticism of so many late Victorian intellectuals. Darwin never became an atheist — towards the end of his life he admitted that sometimes he had flashes of belief — but generally found agnosticism both emotionally and intellectually satisfying. He never wanted to deny his beliefs or non-beliefs, but as an English gentleman never wanted to parade them. His closest personal friend was the local vicar, with whom he worked for the welfare of the villagers and with whom he carried on numerous friendly, if argumentative, discussions over the dinner table. In the end, the State Church had no problem calling Darwin home.

Darwin knew of what we might call the organic origins problem. Indeed, the genesis of species was a hot topic when he was a young man. The astronomer-philosopher John F. W. Herschel, writing to Lyell (in a letter that became public), referred to it as the "mystery of mysteries" [Cannon, 1961]. Darwin also knew of the evolutionary solution. His grandfather Erasmus Darwin, first in his major work *Zoonomia* [1794–1796] and then in verse had held forth about evolution. Robert Grant had almost certainly talked to Darwin on the topic. And then in the second volume of the *Principles*, Lyell had given a very clear exposition of the evolutionism of the French biologist Jean Baptiste de Lamarck [1809]. Poor Lyell was in a quandary. By rights, as a uniformitarian he should have agreed that organisms have natural origins. He certainly argued at length that their demise is natural, and hinted that their origin is likewise. But he could never really bring himself to accept evolution (late in life he more or less staggered over the line). It has been suggested that the progressive nature of evolution went against his steady-state view of the world, but a much more probable cause is the case of *Homo sapiens*. Deist or not, Lyell could not stomach the idea that we might have natural causes. Hence no evolution, even though a whole generation of biologists

(notably including Herbert Spencer) was to be converted to evolution by Lyell's exposition of Lamarck. They read the positive side and ignored the critique.

Darwin did not become an evolutionist on becoming a deist, but the deism primed him. He was looking for laws and their consequences as evidence of God's power. In 1835, having finished in South America, the *Beagle* set out across the Pacific, intending to return home by circling the globe. It put into the Galapagos Archipelago, that group of volcanic islands on the equator (now belonging to Ecuador). As always, Darwin was making collections, especially of the little birds he found there — mockingbirds and finches, of different species and similar to specimens he had found on the South American mainland. He did not realize that the differences were significant until he dined with the governor of the islands. Famously the Galapagos has giant tortoises and the governor told Darwin that one can distinguish the specimens from one island to another. Darwin realized that the same is true also of the birds he had been catching.

This discovery caused a massive problem in natural theology, if nothing else. At Cambridge, Darwin had been much exposed to the argument from design – the argument that God's existence and nature can be inferred from the organization of the world, especially the organic world. The move to deism changed his belief in this not one wit. But think about the Galapagos animals, birds and tortoises. The islands are only a few miles apart, and yet they carry different forms. Darwin had seen cases in South America where the same species of birds might be found in the jungles of Brazil and the snowy deserts of Patagonia. Did God so love the Galapagos that he made different kinds for each island? And why did he make the denizens like South America rather than like Africa, say?

Darwin held fire until he returned to England. Then he showed his bird specimens to John Gould, England's leading ornithologist. (This shows how fast and how far Darwin, thanks to his Cambridge mentors, was rising up the scientific scale.) Gould declared the birds unambiguously to be members of different species. There was only one answer to a young man who saw unbroken law as the key to science and theology: evolution. In the early spring of 1838, Darwin moved across to the belief that species are formed by natural processes, by a long connected chain of development.

Why was the human question never the barrier for Darwin that it was for Lyell? The reason seems to lie in one of those accidents of history. On an earlier voyage, the *Beagle* had picked up three of the natives who live at the bottom of South America, in Tierra del Fuego. The captain (Robert Fitzroy) had taken them back to England, where they were an overnight sensation. Suitably civilized, they were now being returned on Darwin's trip, together with a missionary, to start a process of civilizing their kinfolk and converting them to Christianity. The project was an absolute disaster. The England-returning natives reverted almost overnight to the most abject savage state, and the poor missionary had to be rescued. The ship's naturalist learnt a lesson he never forgot. We are just a fine suit of clothes away from gross animality. Never let us pretend that we are not part of the organic world.

Darwin was not about to tell the scientific world of his conversion. The big problem with evolution in the late 1830s was not so much its anti-biblical cast, but more the fact that it was associated with radical ideas. Evolution was taken to be unabashedly progressive — scum to animal, monad to man, blob to Briton. Indeed, amusingly, Lyell reinforced this connection by portraying Lamarck's evolution as trying to explain a progressive fossil record, when truly Lamarck was doing no such thing at all. The trouble with biological progress was that properly it was taken to be a cousin of social progress, and with reason social progress was taken to be a radical idea — it had led to the American Revolution and, far worse, the French revolution. More than this, the social notion of progress was thought anti-Christian. It implies that we humans unaided can make for a better world, whereas the Christian knows that we are sinners, and only through God's grace is improvement possible. Providence rather than progress.

The late 1830s radicals, including Robert Grant, had taken up evolution with enthusiasm, seeing it as support for their anti-Christian, progressivist program [Desmond, 1989]. There was no way that the socially respectable young Darwin, terrifically ambitious and desperate to be well regarded in the scientific world, was going to blot his copy book by announcing himself an evolutionist. (Interestingly, the one person with whom he might have discussed evolutionary ideas in a semi-friendly manner was a new pal, the anatomist Richard Owen. Although Owen was later to become the violent enemy of the Darwinians, after the *Beagle* voyage he and Darwin became close and there is evidence that there was some vague talk about the transmutation of forms.)

NATURAL SELECTION

So Darwin stayed quiet, filling notebook after notebook with his readings and speculations [Barrett *et al.*, 1987]. Why did he bother? Simply because he was an Englishman, educated at Cambridge. The greatest scientist of them all, another Cambridge-educated Englishman, Isaac Newton had made his name by finding the cause behind the Copernican Revolution. Before Newton, others like Kepler and Galileo had done the spadework. Now it was for Newton to provide the universal force of gravitational attraction, that made everything work. Darwin wanted to be the Newton of biology. He was not the first to propose evolution. He was going to be the first to find the cause. Remember the point about Darwin not being the Christian God and not making everything form nothing. Keep this in mind, as we follow him through to his great discovery.

It seems that it was a quick move for Darwin to sense that some kind of selection must be the key behind change. Darwin knew of so-called Lamarckism – the inheritance of acquired characteristics — it is in *Zoonomia* apart from anything else. He always accepted that it is true in some respects, but he could see that it could never be a fully adequate explanation of change. Apart from anything else, animal and plant breeders told him this. But they did also tell him that if you want change in your cows or sheep or turnips or pigeons, then what you must

do is select rigorously and breed only from the desired forms. Darwin was well placed to pick up on this information. The Darwins lived in rural England and the younger generation of Wedgwoods could afford to play at gentlemen farming, so Charles was able to tap into the lore of the breeders. (Something at that time being much pursued and discussed, given its significance in the Agricultural Revolution, a necessary concomitant to the Industrial Revolution.)

But does it make sense to convert the artificial world of the breeders into a universal force for change in nature? It certainly does. It was here that Darwin's Cambridge, Anglican background came into play. If we can say that his later deism made Darwin an evolutionist, we can say that it was his earlier theism that made him a Darwinian, meaning the man of the mechanism. Darwin just knew that the selection hypothesis had to be on the right track. Again Archdeacon Paley figures in the story. Previously he was important for his contributions to revealed theology, miracles and their meaning. Now, moving from faith to reason, it was natural theology that came into play. Paley's *Natural Theology* [1809], with the detailed discussion of the argument from design, was taken as absolutely basic by Darwin. In the case of the Galapagos animals, it had been the meaning of their distributions that had been at issue. Now it was the very organization of organisms themselves that was important. Paley emphasized that organisms are not just thrown together. They are as if designed — Paley thought because they are designed. Organisms show adaptation, or to use the language that stayed with Darwin all of his life, they show "final cause." The hand and the eye show purpose, function, end, to the cause of helping their possessors. It was just this teleological aspect of the living world to which selection spoke. The breeder of pigs, for instance, wants bigger and fatter pigs, to suit his ends. The size and the fatness are not incidental — they are designed, just like the eye seems to be. Selection is a final-cause-producing mechanism.

But still Darwin could not see how to get selection to work in nature. If anything, the selective powers of breeders seem limited and perhaps transitory. Then at the end of September 1838, Darwin read a very conservative political tract by another Anglican clergyman, the Reverend Thomas Robert Malthus. In his *Principle of Population* (Darwin read the sixth edition of 1828) Malthus argued that food supplies will always be outstripped by population pressure. The latter tend to go up geometrically (1, 2, 4, 8,. . .), whereas the former go up just arithmetically (1, 2, 3, 4,. . .). There will inevitably be struggles for existence, for only a few can survive. This means that grandiose plans for state welfare are doomed to failure. If you feed the poor in this generation, you will only have more in the next. (In person, Malthus was a gentle man, who cared deeply for his parishioners. He saw his calculations as demonstrating God's providence. If there were no stimulus to action, then we would simply spend our days doing nothing. As it is, we had better get up and work for our bread.)

Under pressure, Malthus had conceded that the struggles could be avoided, if only people would practice "prudential restraint" — namely not marrying too early and so forth. (Despite the fact that Aldous Huxley's *Brave New World* refers to

putting in contraceptives as Malthusian drill, Malthus himself was horrified at such practices.) Darwin saw that there can be no prudential restraint in the animal and plant worlds, and hence there would be ongoing unrestricted struggles for survival. This was the motive force that he needed behind a natural form of selection. More organisms are born than can survive and reproduce, some win and some lose, and on average the winners are different from the losers. Given time, this process leads to ongoing change or evolution, and in the direction of adaptive advantage. The winners have things that the losers do not have, and these are the things that make for victory.

An idea is not a theory. Darwin mulled over his ideas, and in 1842 wrote up a 35 page sketch of his thinking followed by a longer, 230 page essay in 1844 [Darwin and Wallace, 1958]. But he did not publish. Instead he then went off for almost ten years on a massive taxonomic study of barnacles [Darwin, 1851a; b; 1854a; b]. Thomas Henry Huxley's explanation for the delay was that Darwin recognized that he was not a trained biologist and so set off to get that training and the authority that came with it. The trouble with this explanation is that there really is not that much difference between these early versions and the account that finally appeared in 1859, the *Origin of Species*. More plausible is the suggestion that in 1844, an anonymous author (since revealed as the Scottish publisher Robert Chambers) published an evolutionary work, *The Natural History of the Vestiges of Creation*. All hell was let loose, with the senior members of Darwin's circles — Sedgwick and Whewell particularly — excoriating it in articles and books. Darwin — the now really sick Darwin — had no taste for that kind of controversy. So he sat on his work, year after year. Finally, in the mid 1850s, with supporters now around him — with the old guard no longer in major authority, with the idea of evolution (thanks in major part to Chambers) no longer so controversial — Darwin started to put together his ideas for publication. A massive work that would convince by argument and footnote.

But he had nearly left things too late. In the summer of 1858, Darwin of all people — surely showing that the scientific world was becoming aware of Darwin's evolutionism — received from a young, naturalist collector in the Malay Archipelago, an essay based on the idea of natural selection. The mechanism was not called that, but there was no ambiguity — even the Malthusian calculation was included. Darwin's friends, Lyell and Hooker, immediately arranged for publication in the *Transactions of the Linnaean Society*, of this essay, by Alfred Russel Wallace, together with selected passages of Darwin's writings. Then, claims to originality and priority established, Darwin set about writing up his ideas. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* appeared at the end of November, 1859. 1250 copies were published and booksellers snapped them up on the first day. Darwin started right in on a second edition. (In all there were six editions, the last appearing in 1872.)

THE "ORIGIN OF SPECIES."

Darwin referred to the *Origin* as "one long argument," and this is the way to read it. For all that it is written in a friendly, almost folksy style, it is a very sophisticated piece of reasoning that reflects the influences that Darwin felt when he was learning to be a scientist in the 1830s. Mention has already been made of Newton. The shadow lay over everything. Darwin set out to find a force, the equivalent in biology of gravitational attraction in physics, and in natural selection he thought he had found it. But one could not, as it were, just plonk the force into the discussion without preparation or argument. The authorities on scientific methodology, primarily Herschel and Whewell, agreed that really one ought to be building a connected system of laws — what in the twentieth century we learned to call hypothetico deductive systems — and that cause must be located within these.

But what kinds of causes? Only the best, what Newton somewhat mysteriously labeled "true causes" or *verae causae*. Now came the question of what constitutes a *vera causa* — what is the mark that one has one? Here Herschel and Whewell parted company. Herschel [1830] took a more empiricist approach, arguing that one must have experience of such causes or analogous phenomena like them. That was a major reason why he liked Lyell's theory of climate. We may not have experienced the past, but we have experienced the Gulf Stream. Whewell [1840] took a more rationalist position. Rather than arguing from experience, he wanted to argue to experience. We may never have encountered a catastrophe or anything like it, but if today's aftereffects point to one, then so be it. He argued that true causes are located at the heart of what he called "consilience of inductions." Unification. Many areas point to a hypothesis; the hypothesis explains many areas.

The debate over geology on one side, this was no mere academic exercise. The wave theory of light was now succeeding beyond all dreams and the methodologists were trying to explain the success. No one could see waves, so why should anyone think them true causes? Herschel [1827] virtually bent himself in half trying to think up analogies — literally, pieces of string and sealing wax — to show how interference patterns could take place. Whewell [1840], rather smugly, sat back and argued that the effects are what counts. Who cares if anyone has seen a light wave or not?

The young Darwin took all of this in, and decided that he had to show his force of natural selection to be a *vera causa*. But empiricist or rationalist? Darwin covered himself in the *Origin* by going both routes! First he introduced the topic of artificial selection, showing just how powerful it could be and how far it is possible to change organisms to one's will.

> The great power of this principle of selection is not hypothetical. It is certain that several of our eminent breeders have, even within a single lifetime, modified to a large extent their breeds of cattle and sheep.... Breeders habitually speak of an animal's organization as something quite plastic, which they can model almost as they please. If I had

> space I could quote numerous passages to this effect from highly competent authorities. Youatt, who was probably better acquainted with the works of agriculturists than almost any other individual, and who was himself a very good judge of animals, speaks of the principle of selection as "that which enables the agriculturist, not only to modify the character of his flock, but to change it altogether. It is the magician's wand, by means of which he may summon into life whatever form and mould he pleases." Lord Somerville, speaking of what breeders have done for sheep, says:
>
> "It would seem as if they had chalked out upon a wall a form perfect in itself, and then had given it existence." [Darwin, 1859, 30–31]

You might think that this is all pretty obvious stuff. Darwin was leading the reader into the main topics gently, by referring to his own route to discovery of natural selection. But the discussion is intended to do more than this, for apart from anything else, general opinion at the time was that the very thing you need to do is to separate natural selection from artificial selection — the latter is very limited and cannot support the former. Wallace, in his essay, argued explicitly that there is no true analogy. Darwin to the contrary pushed the analogy, and again and again through the *Origin* returned to it. For instance, in his discussion of embryology, he wanted to argue that the reason for the similarities of embryos of very different adult forms is that selection does not work on the embryos, only the adults. To make this point, he showed that animal breeders do not work on the young, only the adults, and this shows.

> Some authors who have written on Dogs, maintain that the greyhound and bulldog, though appearing so different, are really varieties most closely allied, and have probably descended from the same wild stock; hence I was curious to see how far their puppies differed from each other: I was told by breeders that they differed just as much as their parents, and this, judging by the eye, seemed almost to be the case; but on actually measuring the old dogs and their six-days old puppies, I found that the puppies had not nearly acquired their full amount of proportional difference. So, again, I was told that the foals of cart and race- horses differed as much as the full-grown animals; and this surprised me greatly, as I think it probable that the difference between these two breeds has been wholly caused by selection under domestication; but having had careful measurements made of the dam and of a three-days old colt of race and heavy cart-horses, I find that the colts have by no means acquired their full amount of proportional difference. (444–445)

Adding: "Fanciers select their horses, dogs, and pigeons, for breeding, when they are nearly grown up: they are indifferent whether the desired qualities and structures have been acquired earlier or later in life, if the full-grown animal possesses them." (446)

The analogy was crucial for Darwin, for it was the empiricist's way of saying that natural selection is a true cause. We have a force, under our control, and this makes reasonable another cause like it in nature. A textbook example of a *vera causa*. In fact, Herschel introduced his discussion of a *vera causa* by asking why we should think that there is a force pulling the moon towards the earth and keeping it in orbit. He asked us to think of a stone on a piece of string, being whirled around in a circle. We sense the force along the string and the need to keep pulling. A force sensed or controlled by us made reasonable the force in nature. Likewise for Darwin.

Darwin was not going to get very formal in his discussion of selection. There would be no explicit laws and tight deductions as the concept was introduced. But he obviously had something like this in mind as he argued to the struggle — the Malthusian calculation is, after all, based on a truth of mathematics.

> A struggle for existence inevitably follows from the high rate at which all organic beings tend to increase. Every being, which during its natural lifetime produces several eggs or seeds, must suffer destruction during some period of its life, and during some season or occasional year, otherwise, on the principle of geometrical increase, its numbers would quickly become so inordinately great that no country could support the product. Hence, as more individuals are produced than can possibly survive, there must in every case be a struggle for existence, either one individual with another of the same species, or with the individuals of distinct species, or with the physical conditions of life. It is the doctrine of Malthus applied with manifold force to the whole animal and vegetable kingdoms; for in this case there can be no artificial increase of food, and no prudential restraint from marriage. [Darwin, 1859, 63]

To get to natural selection, Darwin needed a source of variation — what one might call the raw building blocks of evolution. This he supplied to his satisfaction by surveying the living world and finding that everywhere there is evidence of massive variation in populations. Then he was ready to offer another quasi-formal argument.

> Let it be borne in mind in what an endless number of strange peculiarities our domestic productions, and, in a lesser degree, those under nature, vary; and how strong the hereditary tendency is. Under domestication, it may be truly said that the whole organization becomes in some degree plastic. Let it be borne in mind how infinitely complex and close-fitting are the mutual relations of all organic beings to each other and to their physical conditions of life. Can it, then, be thought improbable, seeing that variations useful to man have undoubtedly occurred, that other variations useful in some way to each being in the great and complex battle of life, should sometimes occur in the course of thousands of generations? If such do occur, can we doubt

> (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind? On the other hand we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable variations and the rejection of injurious variations, I call Natural Selection. (80-81)

Here then was the master mechanism. There was more work to do in elaboration. One important addition was the so-called Principle of Divergence: "from the simple circumstance that the more diversified the descendants from any one species become in structure, constitution, and habits, by so much will they be better enabled to seize on many and widely diversified places in the polity of nature, and so be enabled to increase in numbers" (112). Darwin claimed to have thought of this sometime in the early 1850s, although there are intimations in the earliest species notebooks. For Darwin, the course of evolution was always one of branching. This was built into the Galapagos discovery. A group splitting and the parts going different ways, rather than all change being one of ancestor and descendent without splitting. (Although he admitted some splitting, Lamarck's view of things was essentially that all change is from one form to another, in a line.)

This key principle of divergence was one very much modeled on one of Darwin's favorite metaphors, namely the division of labor — something highlighted in the economic world by the eighteenth-century, Scottish economist Adam Smith, and a foundational principle of industrialists like Darwin's grandfather Josiah Wedgwood. You get much more by breaking things up and specializing — intermediates, the best at nothing, get wiped out quickly. In the organic world, Darwin saw adaptation functioning at its optimum when an organism stuck to its last and did what it could best.

> The advantage of diversification in the inhabitants of the same region is, in fact, the same as that of the physiological division of labor in the organs of the same individual body... No physiologist doubts that a stomach by being adapted to digest vegetable matter alone, or flesh alone, draws most nutriment from these substances. So in the general economy of any land, the more widely and perfectly the animals and plants are diversified for different habits of life, so will a greater number of individuals be capable of there supporting themselves. A set of animals, with their organization but little diversified, could hardly compete with a set more perfectly diversified in structure. (115-116)

Hence, divergence leading to (another famous metaphor) the tree of life.

> The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe this simile largely speaks the truth. The green and budding twigs may represent existing species;

> and those produced during each former year may represent the long succession of extinct species. At each period of growth all the growing twigs have tried to branch out on all sides, and to overtop and kill the surrounding twigs and branches, in the same manner as species and groups of species have tried to overmaster other species in the great battle for life. The limbs divided into great branches, and these into lesser and lesser branches, were themselves once, when the tree was small, budding twigs; and this connexion of the former and present buds, by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups. Of the many twigs which flourished when the tree was a mere bush, only two or three, now grown into great branches, yet survive and bear the other branches; so with the species which lived during long-past geological periods, very few have left living and modified descendants. From the first growth of the tree, many a limb and branch has decayed and dropped off; and these fallen branches of various sizes may represent those whole orders, families, and genera which have now no living representatives, and which are known to us only from being found in a fossil state.... As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever-branching and beautiful ramifications. (129–130)

Darwin still had a little more mop-up work to do. He had but scanty (and generally wrong) ideas about the principles of heredity — what we today call genetics [Vorzimmer, 1970]. He knew that somehow features — new variations — had to be transmitted from one generation to the next. But beyond a hodgepodge of farmers' lore and old wives' tales and fanciers' convictions, he had no real overall picture. One basic problem was that Darwin thought that features would tend to be blended from generation to generation — which of course is true in many respects at the physical level – but at this time he had no underlying theory to explain either the blending or the many exceptions. (A paradox often noted by historians was, at this time, the obscure Moravian monk Gregor Mendel was unweaving the tangle, starting to grasp the true principle of heredity. However, it seems generally agreed that even had Darwin had Mendel's work, he would not necessarily have recognized it for what it was. Mendel knew of Darwin and did not think he (Mendel) had made a breakthrough. When Mendel was rediscovered at the beginning of the twentieth century, it took a decade or more before people really grasped that Mendel and Darwin are complementaries rather than contradictories.)

THE CONSILIENCE

Convincing himself that he had done enough in this respect, Darwin now moved on to the rest of the *Origin* — about three fifths of the total. Here he set about showing that natural selection is a true cause in the rationalist sense — that is, he set about providing a consilience of inductions with natural selection as the unifying causal core. And this he did in fine fashion, as he moved in succession through instinct, paleontology, biogeographical distribution, anatomy, embryology, systematics and more. All of these areas are illuminated and explained by natural selection and in turn they make the truth of selection more secure. And remember, throughout the discussion, the emphasis is not merely on evolution alone, but evolution in the direction of adaptive advantage.

Take by example the honey bee and its hive building. The cells made from wax are perfect hexagons. Why should this be? Because this is the most efficient way of building cells and economizing on wax — a relatively scarce resource and one that requires much labor. Could it be that natural selection is at work here? "We hear from mathematicians that bees have practically solved a recondite problem, and have made their cells of the proper shape to hold the greatest possible amount of honey, with the least possible consumption of precious wax in their construction. It has been remarked that a skilful workman with fitting tools and measures, would find it very difficult to make cells of wax of the true form, though this is effected by a crowd of bees working in a dark hive. Grant whatever instincts you please, it seems as first quite inconceivable how they can make all the necessary angles and planes, or even perceive when they are correctly made" (224). However, showed Darwin in great detail, not only is this possible, it is positively probable. It is all a question of basic instincts and following a few simple rules. The bees are not thinking, they are merrily following the actions as it were preprogrammed into them. Those groups of bees that had the best building patterns survived and reproduced and those that did not did not.

> It seems at first to add to the difficulty of understanding how the cells are made, that a multitude of bees all work together; one bee after working a short time at one cell going to another, so that, as Huber has stated, a score of individuals work even at the commencement of the first cell. I was able practically to show this fact, by covering the edges of the hexagonal walls of a single cell, or the extreme margin of the circumferential rim of a growing comb, with an extremely thin layer of melted vermilion wax; and I invariably found that the colour was most delicately diffused by the bees — as delicately as a painter could have done it with his brush — by atoms of the coloured wax having been taken from the spot on which it had been placed, and worked into the growing edges of the cells all round. The work of construction seems to be a sort of balance struck between many bees, all instinctively standing at the same relative distance from each other, all trying to sweep equal spheres, and then building up, or leaving

> ungnawed, the planes of intersection between these spheres. It was really curious to note in cases of difficulty, as when two pieces of comb met at an angle, how often the bees would pull down and rebuild in different ways the same cell, sometimes recurring to a shape which they had at first rejected. (231-232)

Similar kinds of arguments — partly experimental, partly personal observational, partly general gathering of knowledge from others — occur again and again in this part of the *Origin*. Why do we find a roughly progressive fossil record? Because organisms have changed through time and they have become more and more efficient thanks to the battles leading to selection. This is a kind of forerunner to what today's evolutionists call "arms races" where lines of organisms compete against each other — the prey gets faster and then the predator gets faster — as they refine their adaptations. An addition to a later edition of the *Origin* made the case for humans.

> If we take as the standard of high organisation, the amount of differentiation and specialization of the several organs in each being when adult (and this will include the advancement of the brain for intellectual purposes), natural selection clearly leads towards this standard: for all physiologists admit that the specialization of organs, inasmuch as in this state they perform their functions better, is an advantage to each being; and hence the accumulation of variations tending towards specialisation is within the scope of natural selection" (Darwin 1959, 222).

Biogeography, naturally brought in the Galapagos and related areas.

> Although in oceanic islands the number of kinds of inhabitants is scanty, the proportion of endemic kinds (*i.e.*, those found nowhere else in the world) is often extremely large. If we compare, for instance, the number of the endemic land-shells in Madeira, or of the endemic birds in the Galapagos Archipelago, with the number found on any continent, and then compare the area of the islands with that of the continent, we shall see that this is true. This fact might have been expected on my theory, for, as already explained, species occasionally arriving after long intervals in a new and isolated district, and having to compete with new associates, will be eminently liable to modification, and will often produce groups of modified descendants. (390)

And again:

> The Galapagos archipelago [is] situated under the equator, between 500 and 600 miles from the shores of South America. Here almost every product of the land and water bears the unmistakeable stamp of the American continent. There are twenty-six land birds, and twenty-five of these are ranked by Mr. Gould as distinct species, supposed to

> have been created here; yet the close affinity of most of these birds to American species in every character, in their habits, gestures, and tones of voice, was manifest. So it is with the other animals, and with nearly all the plants, as shown by Dr. Hooker in his admirable memoir on the Flora of this archipelago. The naturalist, looking at the inhabitants of these volcanic islands in the Pacific, distant several hundred miles from the continent, yet feels that he is standing on American land. Why should this be so? why should the species which are supposed to have been created in the Galapagos Archipelago, and nowhere else, bear so plain a stamp of affinity to those created in America? There is nothing in the conditions of life, in the geological nature of the islands, in their height of climate, or in the proportions in which the several classes are associated together, which resembles closely the conditions of the South American coast: in fact there is a considerable dissimilarity in all these respects. On the other hand, there is a considerable degree of resemblance in the volcanic nature of the soil, in climate, height, and size of the islands, between the Galapagos and Cape de Verde Archipelagoes: but what an entire and absolute difference in their inhabitants! The inhabitants of the Cape de Verde Islands are related to those of Africa, like those of the Galapagos to America. (398)

How else can one explain any of these facts except by supposing that they are produced by evolution through natural selection? So Darwin kept going until right at the end, he was able to give his triumphant conclusion.

> It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent upon each other in so complex a manner, have all been produced by laws acting around us. These laws, taken in the largest sense, being Growth with Reproduction; Inheritance which is almost implied by reproduction; Variability from the indirect and direct action of the conditions of life, and from use and disuse; a Ratio of Increase so high as to lead to a Struggle for Life, and as a consequence to Natural Selection, entailing Divergence of Character and the Extinction of less-improved forms. Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved. (489–490)

Some people get terribly excited by the fact that in the *Origin* Darwin never used the word "evolution," and only just squeezes out "evolved" as the last word. This is an emotion based on ignorance. The original sense of the word "evolution" had more to do with embryonic development and only came to mean change through time around the middle of the nineteenth century — Herbert Spencer was much responsible for popularizing this use. People more generally used "transformation" or "transmutation" or, in Darwin's case, "descent with modification." They meant "evolution" in our modern sense [Richards, 1992].

AFTER THE "ORIGIN"

Anyone's life would necessarily be a bit anticlimatical after writing and publishing the *Origin*, and so it proved for Darwin — although he was to live more than twenty more years. He intended to go back to the beginning, writing books for each separate chapter or section of the *Origin*. In the end, Darwin completed only one, because other ideas and projects kept coming to the fore. His first book after the *Origin* was a little work on orchids, intending to show how a topic could be illuminated by someone who took seriously natural selection [Darwin, 1862]. Then later on there were books on climbing plants, on insectivorous plants, and even one on earthworms [Darwin, 1875; 1880; 1881]. A connecting thread for all of these topics was that they could be studied by someone in his own home, someone with the leisure and cash to run simple experiments, someone with a wide range of correspondents who could run around and gather pertinent facts.

The one completed work of the big project was on artificial selection — *The Variation of Animals and Plants under Domestication* [1868]. This was a massive two-volume compendium of just about everything known on the subject. Its main interest for posterity is that Darwin took the opportunity to introduce his "provisional hypothesis of pangenesis." This was his answer to the heredity problem. He argued that all over the body there are little particles, "gemmules," that are carried around and down to the sex cells where they accumulate. Then in reproduction these are passed on to the next generation, the gemmules of parents mingling in the offspring. Note that this hypothesis centrally incorporates a Lamarckian perspective on generation, something to which Darwin always subscribed. A physical change in the adult will be reflected in changes in the gemmules and thus open to transmission across the generations. The hypothesis also assumes that generally features in the offspring will be mixtures of features in the two parents. However, it does leave open the possibility that effects of the gemmules from one parent might swamp the effects of gemmules from the other parent. Moreover, and this was a topic that fascinated Darwin, it lays open the possibility that — since the gemmules themselves remain entire from generation to generation (any blending effects are in the organism and not the causes) — an organism could revert to the features of earlier generations. Some characteristic might be possessed by grandparent say, not the parent, but by the child. The gemmules allowed the possibility of features skipping generations.

Obviously this theory leaves as much unexplained as it explained. Cell theory was now being accepted universally, and Darwin only uneasily used this to underpin his hypothesis. (An earlier unpublished version made no mention of cells.) No real attempt was made to explain how gemmules could switch their effects on and off. Most crucially, no real discussion was given of how the gemmules would be transported around the body and deposited in the sex cells. One would think it would be the blood that would be the key fluid here, but when Darwin's cousin Francis Galton ran experiments shifting blood from one rabbit to another with negative effects, Darwin responded rather huffily that he had never said it was the blood that was important. Which may have been true, but which also leaves the question of transport even more open. Not surprisingly, although Darwin expressed relief at having some hypothesis of any kind, generally it was not a great success or much adopted by others. Although one might legitimately point out that no one else had much idea about heredity. It required years of painstaking labor with the microscope, working out the details of the cell, before answers started to emerge at the beginning of the twentieth century.

The really major project that occupied Darwin in the years after the *Origin* was his work on our species, *Homo sapiens*. His study, *The Descent of Man and Selection with Respect to Sex*, appeared in 1871, with a kind of follow-up volume, *The Expression of the Emotions in Man and in Animals*, in 1872. I noted earlier that Darwin, almost uniquely, never had any problems whatsoever in thinking of our species from a natural perspective. Those Tierra del Fuegians had struck deep into the heart of the Beagle's naturalist. We are animals like the others. Indeed, the very first jottings that we have in Darwin's notebooks, showing that he had grasped the idea of natural selection, deal with humans. And not just any aspect of humans, but that which makes us special.

> [November 27, 1838] An habitual action must some way affect the brain in a manner which can be transmitted.— this is analogous to a blacksmith having children with strong arms.— The other principle of those children. which *chance*? produced with strong arms, outliving the weaker ones, may be applicable to the formation of instincts, independently of habits.— the limits of these two actions either on form or brain very hard to define. [Barrett *et al.*, 1987, N 42-43e]

In the *Origin*, Darwin was circumspect. He knew that the implications of his theory for humans would be the major point of controversy. (He was right about that!) So he wanted first to concentrate on getting the full details of his theory out on the table, as it were. Hence he stayed away from the human question in the *Origin*. But he did not want to conceal his opinions — no one was going to say that Charles Darwin did not realize where it all led. Hence, one of the most understated comments of all time: "Light will be thrown on the origin of man and his history" (488).

The early 1860s saw a spate of books on the man question. Thomas Henry Huxley jumped right in with *Man's Place in Nature* [1863], with a frontispiece

that left no doubt about our continuity with the apes. Charles Lyell also had things to say in *The Antiquity of Man* [1863]. Probably, Darwin would have been happy to let sleeping dogs lie. But it was not to be. It was Alfred Russel Wallace of all people who stirred the pot. As a young man, he seems to have had no more religious belief than anyone. In the 1840s, he was able to read *Vestiges* and at once convert to evolutionism. After the *Origin* was published, and Wallace had returned to England, at first it was he who pushed an evolutionary selection-inspired treatment of the human question in an article that was much praised by Darwin [Wallace, 1864]. But then disaster struck. Wallace became enamored with spiritualism, and started to think that the world is under the control or influence of some great force. Most particularly that the arrival of humans on this planet was not purely natural. And in support of his position, Wallace listed several features of humans that he thought could not possibly have been produced by natural selection. These included things like our hairlessness and our intelligence. Wallace had had first-hand experience of living with natives, and he knew that they had much higher intelligence than most Victorians supposed possible, and yet they do not generally use this intelligence to the full and so clearly it must have arrived through processes other than selection [Wallace, 1870].

Darwin was appalled. Something had to be done, and hence his book on humankind. Much of the work was fairly straightforward and predictable, as Darwin discussed what is known about humans and their origins. Basically, he saw no reason to think that we call for anything but general processes of natural selection. He did pay some attention to actual paths of evolution. In the *Origin*, Darwin had more or less stayed away from these kinds of questions. Although the earlier work on barnacles was not set in an evolutionary context, the knowing reader could with little difficulty work out how Darwin thought that their actual evolution had proceeded. In the *Origin*, Darwin had rather different fish to fry — causes — and so he did not indulge in much path tracing (in modern language, phylogeny tracing). In the *Descent*, Darwin felt able to discuss rival hypotheses and in the end he came down to one favoring an African origin for our species. This tied in with the romantic view that the Victorians then had of the dark (and mysterious) continent. (The *Descent* appeared in the year that H. M. Stanley discovered Dr Livingstone at Ujiji, on the edge of Lake Tanganyika.)

Nevertheless, the *Descent of Man* is far from a conventional book. Taken overall, it has a very peculiar structure and contents. Most of the book is not about humans at all! Right from the beginning back in the early 1840s, Darwin had always argued that there is a secondary form of selection, what he called sexual selection. Influenced by his readings of animal breeders, Darwin saw two kinds of selection — that for general living, as when the breeder makes a bigger cow or a shaggier sheep, and that for competition within the species, as when the breeder makes a stronger bull terrier or a prettier bird. The former was analogous to natural selection and the latter analogous to sexual selection. Refining the concept, the two breeding activities just mentioned led Darwin to distinguish between sexual selection for male combat (the dogs fighting) and sexual selection by female choice

(the two pretty birds displaying before the females).

Sexual selection was introduced into the *Origin*, although it was not given much prominence. Now, in the light of Wallace's apostasy, it was to be brought out into the open and given a job of its own. Darwin wanted to agree with Wallace that natural selection could not account for the peculiarities of humankind, but to disagree with Wallace about the need of a non-natural explanation. Instead, Darwin argued that sexual selection is a major force in human evolution. Hence the need of a general survey of the topic in a book on humans, so that then it could be applied directly to us. As it was.

> With barbarous nations, for instance, the Australians, the women are the constant cause of war both between members of the same tribe and between distinct tribes. So no doubt it was in ancient times; "nam fuit ante Helenam mulier teterrima belli causa." With the North American Indians, the contest is reduced to a system. That excellent observer, Hearne, says:– "It has ever been the custom among these people for the men to wrestle for any woman to whom they are attached; and, of course, the strongest party always carries off the prize. A weak man, unless he be a good hunter, and well-beloved, is seldom permitted to keep a wife that a stronger man thinks worth his notice. This custom prevails throughout all the tribes, and causes a great spirit of emulation among their youth, who are upon all occasions, from their childhood, trying their strength and skill in wrestling." [Darwin, 1871, 2, 324]

Obvious consequences follow:

> There can be little doubt that the greater size and strength of man, in comparison with woman, together with his broader shoulders, more developed muscles, rugged outline of body, his greater courage and pugnacity, are all due in chief part to inheritance from some early male progenitor, who, like the existing anthropoid apes, was thus characterised. These characters will, however, have been preserved or even augmented during the long ages whilst man was still in a barbarous condition, by the strongest and boldest having succeeded best in the general struggle for life, as well as in securing wives, and thus having left a large number of offspring. (2, 325)

PHILOSOPHICAL ISSUES

A matter of some interest, especially to those caring about philosophy, is where Darwin stood on such issues as human knowledge (epistemology) and human morality (ethics). In turning to these questions, it is important to stress that Darwin was always first and foremost a scientist. He was not a philosopher nor was he obsessed with philosophical questions. (Huxley had philosophical pretensions, writing a book on *Hume, With Helps to the Study of Berkeley.*) Darwin had

read Plato as a student (he claimed also to have read Aristotle), as well as more modern philosophers like Locke and Hume. He could talk and write knowledgeably about Kant, knowing the heart of the moral philosophy. ("Duty! Wondrous thought, that workest neither by fond insinuation, flattery, nor by any threat, but merely by holding up thy naked law in the soul, and so extorting for thyself always reverence, if not always obedience; before whom all appetites are dumb, however secretly they rebel; whence thy original?" [Darwin, 1871, 1, 70]). Darwin realized that in writing about humans, he was going to steer close to philosophical issues, and did not shirk the task. However, as a scientist, he was more interested in why we think and act as we do — questions that today would occupy the "evolutionary psychologist" – rather than in the foundations of why we think and act as we do — questions of the philosopher.

As far as reason and knowledge is concerned, Darwin was straightforwardly naturalistic. The reasoning ability is a good thing to have in the struggle for existence, and this is shown by the fact that other animals have it also. There is no reason to think it a special feature of humankind.

> Of all the faculties of the human mind, it will, I presume, be admitted that *Reason* stands at the summit. Few persons any longer dispute that animals possess some power of reasoning. Animals may constantly be seen to pause, deliberate, and resolve. It is a significant fact, that the more the habits of any particular animal are studied by a naturalist, the more he attributes to reason and the less to unlearnt instincts. (1, 46)

Reason and knowledge help animals to survive and reproduce. The same is true of humans. No more discussion of the main point is needed, although Darwin did offer detailed discussion of language, beauty appreciation, beliefs in divinities, to show how these too can be explained in a naturalistic fashion.

If the philosopher persists in asking his or her questions, then he or she might use Darwin's empirical claims as grist for the mill. But there is not much more. Today, one particularly favoured application of evolution to epistemology comes by analogy from the processes of evolution to the processes of knowledge acquirement and development [Ruse, 1986]. People like Stephen Toulmin [1972] and Karl Popper [1974] have argued that just as we get a struggle for existence in the animal world, so also we get a struggle for existence in the intellectual world — one theory beating out another in the competition for acceptance. There are elements of this line of thinking in Huxley, but little in Darwin. The other favoured application of evolution to epistemology comes through taking things literally, arguing that the truths of logic and methodology and so forth are no more (or less) than rules that have proven their worth in the struggle. This approach welcomes Darwin's discussion. Somewhat tantalizingly, in his early notebooks, Darwin himself even flies a kite for this kind of philosophy. "Plato... says in Phaedo that our '*imaginary ideas*' arise from preexistence of the soul, are not derivable from experience. — read monkeys for preexistence" [Barrett *et al.*, 1987, M. 128]. It is no more

than a comment, although it has certainly been enough for some contemporary philosophers to claim Darwin for their own.

There is a detailed discussion of morality in the *Descent*. Darwin is concerned first to get moral norms, and then to see why they have their force, something he puts down to conscience. The precise nature of the moral norms is not of great interest to Darwin — basically he wants something along the lines of common sense, a sort of mishmash of utilitarianism, Kantianism and subscription to the Golden rule or the love commandment. The main thing is that we do have such a sense of obligation.

> I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important. This sense, as Mackintosh remarks, "has a rightful supremacy over every other principle of human action;" it is summed up in that short but imperious word *ought*, so full of high significance. It is the most noble of all the attributes of man, leading him without a moment's hesitation to risk his life for that of a fellow creature; or after due deliberation, impelled simply by the deep feeling of right or duty, to sacrifice it in some great cause. [Then follows the above-given quotation from Kant.] (1, 70)

But how does morality get enforced? Conscience kicks in. "The following proposition seems to me in a high degree probable –namely, that any animal whatever, endowed with well-marked social instincts, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or nearly as well developed, as in man" (1, 71–72). In more detail:

> At the moment of action, man will no doubt be apt to follow the stronger impulse; and though this may occasionally prompt him to the noblest deeds, it will more commonly lead him to gratify his own desires at the expense of other men. But after their gratification, when past and weaker impressions are contrasted with the ever-enduring social instincts, retribution will surely come. Man will feel dissatisfied with himself and will resolve with more or less force to act differently for the future. This is conscience; for conscience looks backwards and judges past actions, inducing that kind of dissatisfaction, which if weak we call regret, and if severe remorse. (1, 91)

There is an interesting subtext here, one that connects directly with a topic much discussed today [Sterelny and Griffiths, 1999]. Charles Darwin was always interested in the level at which natural selection operates [Ruse, 1980]. We have seen that he had no proper knowledge of (what we call) genetics, so he could not take what Richard Dawkins [1976] has called the "selfish gene" approach, but he could and did take (what we call) a hard-line "individual selection" approach. That is to say, he thought that natural selection pits individual against individual, rather than group against group. Any group contests resolve themselves ultimately as

individual contests. Of course, the struggle for existence may take place between an individual and nature, rather than between two individuals, but the end result is that one individual is going to do better against nature than another individual. What you are not going to get therefore is one individual helping another individual just for the sake of niceness. Natural selection will put an end to that pretty sharply, because the selfish individual will take advantage of the nice individual and hence do better in life's struggles.

There are some obvious counter examples and Darwin wrestled with these. Most obviously there are the social insects, a major topic of discussion in the section on instinct in the *Origin*. Why do you get sterile workers giving their all for the nest? Darwin could see how features possessed by the sterile workers could be passed on. Steers have features much prized by breeders, who go back to the family stock to repeat them. But what is the benefit for the workers? Eventually, in the *Origin*, Darwin decided to treat the whole nest as a unit, as an individual, and hence argued that selection could modify parts – just as an ordinary individual has parts (heart, lungs, etc) that can be selected for and against.

But he went on worrying about this issue, and in the 1860s he and Wallace discussed the issue at length. Wallace was always a committed socialist and for him a group perspective was not only scientifically sound but psychologically compelling. (Did Darwin, the grandchild of a major industrialist, feel the same way about individual selection? Possibly.) The two mean differed particularly over sterility. Is the mule sterile because it is good neither as horse or donkey, and this benefits the parent species? Or is the mule sterile by chance? In which case, any benefit for the parent species is quite incidental. Darwin thought chance, because otherwise the mule would be laying down its reproductive life for others, whereas Wallace thought in group terms and the benefit of the parent species. The two men had to agree to disagree. (Wallace, incidentally, other than for humans where he had some very odd ideas about what makes a young man attractive in the eyes of a young woman, was never that enthused about sexual selection by female choice. Given that sexual selection is strongly tilted towards individual selection, this fits the pattern.)

When it came to human morality, however, Darwin rather buckled at the knees. Perhaps here uniquely we have a case where a form of group selection takes over. A tribe where individuals help each other would outcompete a tribe where individuals are selfish. Generosity helps the group even if the individual suffers. But Darwin was not enthusiastic about this move. And after all, tribes were usually kin structures, so we have a kind of family situation as in the social insects. Moreover, in an anticipation of what today's evolutionists call "reciprocal altruism," Darwin also suggested that morality may be a function of "you scratch my back and I will scratch yours."

> But it may be asked, how within the limits of the same tribe did a large number of members first become endowed with these social and moral qualities, and how was the standard of excellence raised?...

> In the first place, as the reasoning powers and foresight of the members became improved, each man would soon learn from experience that if he aided his fellow-men, he would commonly receive aid in return. From this low motive he might acquire the habit of aiding his fellows; and the habit of performing benevolent actions certainly strengthens the feeling of sympathy, which gives the first impulse to benevolent actions. (1, 163-164)

Again the philosopher will ask his or her questions and again the philosopher will have to be satisfied with less than full answers. The most popular way of getting ethics from evolution is by arguing by analogy that, as go the processes of evolution, so go the norms of proper behavior. This is so-called Social Darwinism, and often it takes the form of moving from a bloody struggle for survival in nature to the claim that that is what we have — and what we should have — in society. Unrestricted *laissez faire*, with widows and children going to the wall and the strong surviving and flourishing. Expectedly, most Social Darwinians have been a bit more sophisticated than this [Richards, 1987; Ruse, 2000; 2005]. But there are certainly elements of this kind of thinking in the *Descent of Man*.

> With savages, the weak in body or mind are soon eliminated; and those that survive commonly exhibit a vigorous state of health. We civilised men, on the other hand, do our utmost to check the process of elimination; we build asylums for the imbecile, the maimed, and the sick; we institute poor-laws; and our medical men exert their utmost skill to save the life of every one to the last moment. There is reason to believe that vaccination has preserved thousands, who from a weak constitution would formerly have succumbed to small-pox. Thus the weak members of civilised societies propagate their kind. No one who has attended to the breeding of domestic animals will doubt that this must be highly injurious to the race of man. It is surprising how soon a want of care, or care wrongly directed, leads to the degeneration of a domestic race; but excepting in the case of man itself, hardly any one is so ignorant as to allow his worst animals to breed. (1, 168)

The comments on the virtues of capitalism, on women, on savages, and the Irish rather make one's hair stand on end.

> Or as Mr. Greg puts the case: "The careless, squalid, unaspiring Irishman multiplies like rabbits: the frugal, foreseeing, self-respecting, ambitious Scot, stern in his morality, spiritual in his faith, sagacious and disciplined in his intelligence, passes his best years in struggle and in celibacy, marries late, and leaves few behind him. Given a land originally peopled by a thousand Saxons and a thousand Celts — and in a dozen generations five-sixths of the population would be Celts, but five-sixths of the property, of the power, of the intellect, would belong to the one-sixth of Saxons that remained. In the eternal 'struggle for

> existence,' it would be the inferior and less favoured race that prevailed — and prevailed by virtue not of its good qualities but of its faults." (1, 174)

Fortunately selection is a little more sophisticated. The Irish might breed like rabbits, but they tend to be careless about the raising of their offspring, so many die without issue. The Scots to the contrary, work hard and make sure that their offspring do succeed them. In the long run, nature values the good and decent.

Social Darwinians justify their appeal to the processes of nature by claiming that evolution is progressive. It is right and proper to support natural selection, to cherish it as a beneficent mechanism, because the end justifies the means. Humans are higher than slugs and worms. Some humans (the Scots) are higher than other humans (the Irish). It may seem cruel to ignore the pleas of widows and orphans, but better overall for society if you do. Darwin was certainly a progressionist, about society and about evolution. He realized that in science it was politic to play down explicit progressionism — in the 1830s when he entered the game, (as we have seen) social progress too strongly identified with atheism and radicalism, and biological progress was generally identified with the worst excesses of German romanticism, *Naturphilosophie*, the English then regarding continental philosophy in much the way that they regard it today.

In the first edition of the *Origin*, progress certainly gets mentioned and in the fossil record gets a somewhat tepid endorsement.

> The inhabitants of each successive period in the world's history have beaten their predecessors in the race for life, and are, in so far, higher in the scale of nature; and this may account for that vague yet ill-defined sentiment, felt my many paleontologists, that organisation on the whole has progressed. [Darwin, 1859, 345]

And again:

> As all the living forms of life are the lineal descendants of those which lived long before the Silurian epoch, we may feel certain that the ordinary succession by generation has never once been broken, and that no cataclysm has desolated the whole world. Hence we may look with some confidence to a secure future of great length. And as natural selection works solely by and for the good of each being, all corporeal and mental endowments will tend to progress towards perfection. (489)

Later editions of the *Origin*, for reasons to be given shortly, were more progress-explicit and friendly than earlier editions.

> Although we have no good evidence of the existence in organic beings of an innate tendency towards progressive development, yet this necessarily follows,... through the continued action of natural selection. For the best definition which has ever been given of a high standard of organisation is the degree to which the parts have been specialised

> or differentiated; and natural selection tends towards this end, inasmuch as the parts are thus enabled to perform their functions more efficiently. (Darwin 1959, 241; from the sixth edition of 1872)

The *Descent of Man* is unambiguously progressive, both through the animal kingdom and when we come to our own species. Although Darwin was strongly against slavery — before and during the Civil War — there was no doubt in his mind that white people are superior to all others, and little doubt that the inhabitants of a small island off the northwest coast of Europe are the top among whites. We have seen already Darwin's endorsement of Greg on the Irish and the Scots. So we can certainly say that the ingredients for a Social Darwinian attack on foundations are there in Darwin's writings. But as I have said, he was not really a philosopher and we do not find explicit inferences being drawn. Nor do we find any other attempt to make a philosophical picture from his science. Darwin was an evolutionary psychologist, first and foremost.

THE DARWINIAN REVOLUTION

This now brings to an end the direct examination of Darwin and his work. Let us conclude by turning to broader issues. Darwin was not a man working in isolation. He was influenced by people who came before him, and obviously he influenced those who came after him. General opinion is that he was one of the great thinkers of all time — indeed it has been said that there is no greater discovery than natural selection. The Darwinian Revolution is one of those events in human history that stands right up there, with the Scientific Revolution in thought and with the American, French, and Russian Revolutions in politics. It is about this revolution we must now turn our attention.

Let us not waste time asking if there was indeed a Darwinian Revolution. Some nervous historians have started to worry whether such talk is ever appropriate, so let the robust philosopher in us brush past this anxiety (see [Hodge, 2005]). By any significance measure, the Darwinian Revolution was revolutionary. At the beginning of the nineteenth century, by and large people did not believe in evolution. At the end of the nineteenth century, by and large people did believe in evolution. More than this, they accepted that it applies to our own species, *Homo sapiens*. This was a terrific move. We can agree that it was not necessarily purely a scientific revolution. Perhaps it was not even primarily a scientific revolution, being more one to do with religion — Does God still exist and what does He care about us? — or culture or whatever. But it was a revolution and moreover, whether the most important overall factor or not, science was a very important factor. We can go further and say that it was the prime causal factor, for without the scientists I do not see how you could have had a shift to what is (after all) a scientific claim: organisms, including humans, evolved.

What about a more serious worry, that Darwin gets too much credit for the revolution? How much credit does Charles Robert Darwin merit for the revolution

that carries his name? In one sense, no one can deny that he deserves some, a lot in fact. Before the *Origin of Species* appeared in 1859, the idea of evolution was a minority position and in many respects not very respectable. After the *Origin*, it became in many circles — middle class and working class, religious and not — the accepted position on origins. More than this, Darwin put forward the mechanism of natural selection, and today this is generally accepted as the right mechanism. Darwin got it right about causes.

But there is more to the question than this. Start with the period before Darwin. There was already significant acceptance of evolutionary ideas. In Germany, the despised *Naturphilosophen* were inclined to evolution. Even Goethe, towards the end of his long life, embraced the idea [Richards, 2003]. In France, there was a whole group of evolutionists around Lamarck [Corsi, 1988; 2005]. And this continued through the century. Britain too yields many evolutionists, starting with Charles Darwin's own grandfather, Erasmus. He did not merely influence his grandson. *Zoonomia* was translated into German and read (and commented on) by the aged Immanuel Kant [Ruse, 2006]. Add to this Robert Grant and Robert Chambers and the *Vestiges*. For all of its controversial nature and its being hated by Sedgwick and Whewell and company, this latter work was a major inspiration for many, and not just Wallace. The poet Alfred Tennyson, as he struggled to finish (what rapidly became) his much-loved and read poem, *In Memoriam,* drew heavily on Chambers's thinking. Apparently Tennyson's long-dead friend Arthur Hallam, in whose memory the poem was written, was an anticipation of a higher form, too precious for life today.

> A soul shall strike from out the vast
> And strike his being into bounds,
>
> And moved thro' life of lower phase,
> Result in man, be born and think,
> And act and love, a closer link
> Betwixt us and the crowning race...
>
> Whereof the man, that with me trod
> This planet, was a noble type
> Appearing ere the times were ripe,
> That friend of mine who lives in God.

Finally, mention the general man of letters and science, Herbert Spencer, in the 1850s just beginning his dizzying rise upwards as the people's philosopher in Britain and the rest of the world. In the decade before Darwin, he was publishing evolutionary ideas including a clear statement of natural selection [Ruse, 1996; Richards, 1987].

So there can be no claim that Darwin was the first evolutionist or even the first with natural selection. (Do not forget Alfred Russel Wallace either.). Yet, having said all of this, Darwin's reputation emerges intact. Before Darwin evolution was still little more than a pseudo science, generally accepted because people liked the

progressivist sentiments at its heart. Darwin made the idea of evolution serious. Thanks to him, the idea was not just plausible but, for most folks, absolutely compelling. The way he tied everything together in a consilience was definitive — back then and now. Darwin did this in the *Origin*. The same is true of natural selection. Darwin may have borrowed from others but it was he who made something of it all. Any fool can take pigments and paint a picture of flowers. It took Van Gogh to paint the sunflowers. It took Darwin to write the *Origin*.

This now takes us to the time after the *Origin*. Darwin put together the idea of evolution and made it compelling, and it was because of him (together with his various supporters) that people were converted to the idea of evolution. However, after the *Origin*, it is well known that natural selection was a flop. No one took it up, and it languished until the 1930s, when the population geneticists like Ronald A. Fisher [1930] in Britain and Sewall Wright [1931; 1932] in America melded Darwinian selection with Mendelian genetics to make the new theory, so called neo-Darwinism or the synthetic theory of evolution. Hence in major respects, a critic might complain that, even if you grant that there was a revolution, it was not very Darwinian. Indeed the one historian (Peter Bowler) has gone so far as to write a book with the title, *The non-Darwinian Revolution*!

There is truth in this charge, if one is concentrating less on the fact of evolution and more on the mechanism of natural selection. Scientists generally did not pick it up and use it. There was not a new field of selection studies. Partly, this non-development was scientific. There were perceived problems with selection. Without a good theory of heredity, no one could see how the effects of selection could be long lasting [Vorzimmer, 1970]. Also the age of the earth was a problem. Not knowing about radio-active decay and its warming effects, physicists thought that the earth is much younger than it really is, and it seemed that a leisurely process like selection would never get the job done in time [Burchfield, 1975]. Partly, there were other factors for the non-development, some in science and some outside. Someone like Huxley was never really that interested in adaptation and design, so for him selection was not really needed anyway [Desmond, 1994; 1997]. Evolution was what he needed. Outside science, the religious were happy to accept evolution, but they still wanted a bit of guidance to get organisms, especially humans. So they were into directed mutations and so forth, and eschewed the full implications of the blind, cruel process of selection.

However, this is only half the story. Even if we take into account the scientific problems with natural selection, there was something really odd going on. Notwithstanding the great enthusiasm for evolution as such, the years after the *Origin*, measured by the standards of mature professional science, were generally speaking an absolute disaster for evolutionary studies. It was not so much that people did not like selection and looked for other causes — although some did — but that people were really not that interested in causes at all. They truly were not that interested in evolution as a tool of science. What they sought, and found, was a basis for a kind of secular humanism, a Christianity substitute that could be used as the ideology of the new society that they were determined to build [Ruse,

1996; 2005].

People like Thomas Henry Huxley were determined to reform Victorian Britain, as were others in America after the Civil War. They worked long and hard (and successfully) at transforming and upgrading primary, secondary and tertiary education, the military, the civil service, the medical profession, and more. They saw the established church as a major ally of those whom they were fighting — the aristocrats, the landed gentry, and others with vested interests in opposing change. Hence, they sought their own secular equivalent and in a progressivist popular evolutionism they found it. They did not want a tool of research, but they did want something that could give a moral message. And again, in Social Darwinism they found it. (This is why one should be wary of assuming that all Social Darwinians subscribed to a simple theory of *laissez faire*. In fact, like the Christians they were trying to replace, Social Darwinians subscribed to moral views of many different kinds, and there were strong internal disagreements. See [Ruse, 2000; 2005].)

It is true that there was some professional evolutionary work, but it tended to the decidedly second-rate — phylogeny tracing — and it was increasingly out of touch with reality. And it tended to be brushed aside as Huxley, the Saint Paul of the Darwinism movement (an analogy drawn by his contemporaries), preached the gospel of popular Darwinism non-stop — at working men's clubs, in the newspapers, as the president of numerous societies, and so forth. Evolution became a Christianity substitute, a secular religion, to promote the kind of society that Huxley and his chums wanted to create. There was no wonder that Herbert Spencer, who was right into this sort of thing, was more influential than Darwin. There was no wonder that Darwin — a man who had genuinely hoped to see a science of evolution studies with natural selection at its core — realized that it was better to have half a cake rather than none at all. And so he went along with the flow. After all, he did believe in most of the ideology anyway. As noted, there was much more about progress in later editions of the *Origin*, and the *Descent* throughout adopts a relaxed, user-friendly style. Darwin may not have identified explicitly with Social Darwinism and the related beliefs, but as we have seen he was no foe of the movement.

So, in a way, the critic is right. If you are talking about pure science, in this sense the late nineteenth century saw only a very limited Darwinian Revolution. But, putting things in long-term context, let us never forget that — whether or not his contemporaries got enthused about natural selection — Darwin did get it right! Natural selection is the main cause of evolutionary change. It took seventy-five years for this to become apparent, but then it did and it has stayed that way. So, if you are prepared to use the present as a guide to the past, so long as you do not gloss over how history took time to develop, there is no reason to deny Darwin's role in the Darwinian Revolution. And much good reason to think that the revolution is appropriately named.

Finally, let us ask about the nature of the revolution, meaning more about what kind of revolution it really was. Was there a change of paradigms in the sense described by Thomas Kuhn in his *The Structure of Scientific Revolutions*? Was

there a switch of world views — perhaps even a switch of worlds — that required more of a leap of faith than an appeal to reason? Or was the change smoother, as more conventional philosophies of science might lead one to expect. Was the change more (say) in a Popperian vein, where basically the facts told against the older position and people shifted because this was the reasonable thing to do? The prize goes to both disjuncts! At a broad level, there are certainly Kuhnian aspects to the revolution. Most strikingly, there were people who simply could not see the other side's point of view — clever people, that is, who knew the ins and outs of the issues. Most prominent of these was Louis Agassiz [E. C. Agassiz, 1889]. He had staked out an idealistic position before the *Origin* [Agassiz, 1859] — one that came directly from his *Naturphilosophen* teachers (Friedrich Schelling and Lorenz Oken) when he was a student in Munich — and try as he might, he could never accept evolution.

Something like this makes perfect sense on the Kuhnian scenario and fits uncomfortably into a philosophy of science that makes rational choice the sole criterion of theory change. However, in other respects the Darwinian Revolution seems very non-Kuhnian. People may not have been able to see the viewpoints of others, but it is hard to say that this was a function of different facts, a key element in Kuhn's theory of change. Fact-change is simply not true of the Darwinian Revolution. Everything about Darwin himself denies this claim. We have seen this again and again. Darwin was like Plato's Demiurge, shaping what he already had. This applied to ideas as well as facts. Everyone knew about Malthus, for instance, but it was Darwin's genius to put the ideas into a theory of change rather than a theory that argued that change is impossible. Likewise, the facts of the successes of animal and plant breeders were well known. It was again Darwin's genius to make something of these facts. The same is true of so much else — the vaguely progressive fossil record, for instance, and the peculiarities of biogeography. Particularly important for Darwin were earlier discoveries in embryology. Darwin seized on the similarities of embryos and made this a key support for the arguments of the *Origin*.

So, we have to say that the Darwinian Revolution does not fit readily into simple theories of theory change. Probably in part this was because it was not simply a scientific revolution, but one that included religious beliefs and thoughts about the status of humans and (particularly for people like Huxley, including many today) hopes for the kind of secular, progressivist society that one thinks is the ideal for humankind today. I return to the metaphor of a kaleidoscope. It was not so much that there was a change of facts, but that the facts were shaken up and pushed into making a whole new picture — a picture that some welcomed, some disliked, but towards which few were indifferent. And that thought is no bad way to end this discussion of Charles Darwin and the revolution with which his name is associated.

BIBLIOGRAPHY

[Agassiz, 1885] E. C. Agassiz, ed. *Louis Agassiz: His Life and Correspondence.* Boston: Houghton Mifflin, 1885.

[Agassiz, 1859] L. Agassiz. *Essay on Classification* . London: Longman, Brown, Green, Longmans, and Roberts and Trubner, 1859.

[Barrett *et al.*, 1987] P. H. Barrett, P. J. Gautrey, S. Herbert, D. Kohn, and S. Smith, ed. *Charles Darwin's Notebooks, 1836-1844.* Ithaca, N.Y.: Cornell University Press, 1987.

[Bowler, 1988] P. J. Bowler. *The non-Darwinian Revolution: Reinterpreting a Historical Myth.* Baltimore, Md.: Johns Hopkins University Press, 1988.

[Browne, 1995] J. Browne. *Charles Darwin: Voyaging. Volume 1 of a Biography.* New York: Knopf, 1995.

[Browne, 2002] J. Browne. *Charles Darwin: The Power of Place. Volume II of a Biography.* New York: Knopf, 2002.

[Burchfield, 1975] J. D. Burchfield. *Lord Kelvin and the Age of the Earth.* New York, N.Y.: Science History Publications, 1975.

[Cannon, 1961] W. F. Cannon. The impact of uniformitarianism. Two letters from John Herschel to Charles Lyell, 1836-1837. *Proceedings of the American Philosophical Society* 105: 301-14, 1961.

[Chambers, 1844] R. Chambers. *Vestiges of the Natural History of Creation.* London: Churchill, 1844.

[Corsi, 1988] P. Corsi. *The Age of Lamarck.* Berkeley : University of California Press, 1988.

[Corsi, 2005] P. Corsi. Before Darwin: Transformist concepts in European natural history. *Journal of the History of Biology* 38, 2005.

[Cuvier, [1813]1822] G. Cuvier. *Theory of the Earth.* 4th ed. Edited by Robert Jameson. Edinburgh: William Blackwood, [1813]1822..

[Darwin, 1839] C. Darwin. Observations on the parallel roads of Glen Roy, and of other parts of Lochaber in Scotland, with an attempt to prove that they are of marine origin. *Philosophical Transactions of the Royal Society of London*, no. 39-82, 1839.

[Darwin, 1842] C. Darwin. *The Structure and Distribution of Coral Reefs.* London: Smith Elder, 1842.

[Darwin, 1845] C. Darwin. *Journal of Researches into the Natural History and Geology of the Countries Visited during the Voyage of H.M.S. Beagle round the World [2nd ed.].* London: John Murray, 1845.

[Darwin, 1851a] C. Darwin. *A Monograph of the Fossil Lepadidae; or, Pedunculated Cirripedes of Great Britain.* London: Palaeontographical Society, 1851.

[Darwin, 1851b] C. Darwin. *A Monograph of the Sub-Class Cirripedia, with Figures of all the Species. The Lepadidae; or Pedunculated Cirripedes.* London: Ray Society, 1851.

[Darwin, 1854a] C. Darwin. *A Monograph of the Fossil Balanidae and Verrucidae of Great Britain.* London: Palaeontographical Society, 1854.

[Darwin, 1854b] C. Darwin. *A Monograph of the Sub-Class Cirripedia, with Figures of all the Species. The Balanidge (or Sessile Cirripedes); the Verrucidae, and C.* London: Ray Society, 1854.

[Darwin, 1859] C. Darwin. *On the Origin of Species.* London: John Murray, 1859.

[Darwin, 1862] C. Darwin. *On the Various Contrivances by which British and Foreign Orchids are Fertilized by Insects, and On the Good Effects of Intercrossing.* London: John Murray, 1862.

[Darwin, 1868] C. Darwin. *The Variation of Animals and Plants Under Domestication.* London: Murray, 1868.

[Darwin, 1871] C. Darwin. *The Descent of Man.* London: John Murray, 1871.

[Darwin, 1872] C. Darwin. *The Expression of the Emotions in Man and Animals.* London: John Murray, 1872.

[Darwin, 1875] C. Darwin. *Insectivorous Plants.* London: John Murray, 1875.

[Darwin, 1880] C. Darwin. *The Power of Movement in Plants.* London: John Murray, 1880.

[Darwin, 1881] C. Darwin. *The Formation of Vegetable Mould, Through the Action of Worms, with Observations on their Habits.* London: Murray, 1881.

[Darwin, 1959] C. Darwin. *The Origin of Species by Charles Darwin: A Variorum Text.* Editor M Peckham. Philadelphia, Pa.: University of Pennsylvania Press, 1959.

[Darwin, 1969] C. Darwin. *Autobiography*. New York : Norton, 1969.
[Darwin and Wallace, 1958] C. Darwin and A. R. Wallace. *Evolution by Natural Selection*. Foreword by Gavin de Beer. Cambridge: Cambridge University Press, 1958.
[Darwin, [1794-1796]1801] E. Darwin. *Zoonomia; or, The Laws of Organic Life*. 3rd ed. London: J. Johnson, [1794-1796]1801.
[Dawkins, 1976] R. Dawkins. *The Selfish Gene*. Oxford: Oxford University Press, 1976.
[Desmond, 1989] A. Desmond. *The Politics of Evolution: Morphology, Medicine and Reform in Radical London*. Chicago: University of Chicago Press, 1989.
[Desmond, 1994] A. Desmond. *Huxley, the Devil's Disciple*. London: Michael Joseph, 1994.
[Desmond, 1997] A. Desmond. *Huxley, Evolution's High Priest*. London: Michael Joseph, 1997.
[Desmond and Moore, 1992] A. Desmond and J. Moore. *Darwin: The Life of a Tormented Evolutionist*. New York: Warner, 1992.
[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press, 1930.
[Herbert, 2005] S. Herbert. *Charles Darwin, Geologist*. Ithaca, N.Y.: Cornell University Press, 2005.
[Herschel, 1827] J. F. W. Herschel. Light. In *Encylopaedia Metropolitana*. Edited by E Smedley, and et al. London: J. Griffin, 1827.
[Herschel, 1830] J. F. W. Herschel. *Preliminary Discourse on the Study of Natural Philosophy*. London: Longman, Rees, Orme, Brown, Green, and Longman, 1830.
[Hodge, 2005] M. J. S. Hodge. Against "Revolution" and "Evolution". *Journal of the History of Biology* 28, 2005.
[Huxley, 1863] T. H. Huxley. *Evidence as to Man's Place in Nature*. London: Williams and Norgate, 1863.
[Huxley, 1879] T. H. Huxley. *Hume*. London: Macmillan, 1879.
[Kuhn, 1962] T. Kuhn. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.
[Lamarck, 1809] J. B. Lamarck. *Philosophie zoologique*. Paris: Dentu, 1809.
[Lyell, 1830-1833] C. Lyell. *Principles of Geology: Being an Attempt to Explain the Former Changes in the Earth's Surface by Reference to Causes now in Operation*. London: John Murray, 1830-1833.
[Lyell, 1863] C. Lyell. *The Antiquity of Man*. London: Murray, 1863.
[Malthus, [1826]1914] T. R. Malthus. *An Essay on the Principle of Population (Sixth Edition)*. London: Everyman, [1826]1914.
[Paley, [1794]1819] W. Paley. *Evidences of Christianity (Collected Works: III)*. London: Rivington, [1794]1819.
[Paley, [1802]1819] W. Paley. *Natural Theology (Collected Works: IV)*. London: Rivington, [1802]1819.
[Popper, 1972] K. R. Popper. *Objective Knowledge*. Oxford : Oxford University Press, 1972.
[Richards, 1987] R. J. Richards. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press, 1987.
[Richards, 1992] R. J. Richards. *The Meaning of Evolution: The Morphological Construction and Ideological Reconstruction of Darwin's Theory*. Chicago: University of Chicago Press, 1992.
[Richards, 2003] R. J. Richards. *The Romantic Conception of Life: Science and Philosophy in the Age of Goethe*. Chicago: University of Chicago Press, 2003.
[Rudwick, 1974] M. J. S. Rudwick. Darwin and Glen Roy: a 'great failure' in scientific method? *Studies in the History and Philosophy of Science* 5: 97-185, 1974.
[Ruse, 1979] M. Ruse. *The Darwinian Revolution: Science Red in Tooth and Claw*. Chicago: University of Chicago Press, 1979.
[Ruse, 1980] M. Ruse. Charles Darwin and group selection. *Annals of Science* 37: 615-30, 1980.
[Ruse, 1986] M. Ruse. *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*. Oxford: Blackwell, 1986.
[Ruse, 1996] M. Ruse. *Monad to Man: The Concept of Progress in Evolutionary Biology*. Cambridge, Mass.: Harvard University Press, 1996.
[Ruse, 1999a] M. Ruse. *The Darwinian Revolution: Science Red in Tooth and Claw*. second ed. Chicago: University of Chicago Press, 1999.
[Ruse, 1999b] M. Ruse. *Mystery of Mysteries: Is Evolution a Social Construction?* Cambridge, Mass.: Harvard University Press, 1999.

[Ruse, 2000] M. Ruse. *The Evolution Wars: A Guide to the Controversies*. Santa Barbara, California: ABC-CLIO, 2000.
[Ruse, 2003] M. Ruse. *Darwin and Design: Does Evolution have a Purpose?* Cambridge, Mass.: Harvard University Press, 2003.
[Ruse, 2005] M. Ruse. *The Evolution-Creation Struggle*. Cambridge, Mass.: Harvard University Press, 2005.
[Ruse, 2006] M. Ruse. Kant and evolution. *Theories of Generation.* editor J Smith. Cambridge: University of Cambridge Press, 2006.
[Sterelny and Griffiths, 1999] K. Sterelny and P. E. Griffiths. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press, 1999.
[Toulmin, 1972] S. Toulmin. *Human Understanding*. Oxford : Clarendon Press, 1972.
[Vorzimmer, 1970] P. J. Vorzimmer. *Charles Darwin: the Years of Controversy*. Philadelphia, Penn.: Temple University Press, 1970.
[Wallace, 1864] A. R. Wallace. The origin of human races and the antiquity of man deduced from the theory of natural selection. *Journal of the Anthropological Society of London* 2: clvii-clxxxvii, 1864.
[Wallace, 1870] A. R. Wallace. *Contributions to the Theory of Natural Selection: A Series of Essays*. London: Macmillan, 1870.
[Whewell, 1837] W. Whewell. *The History of the Inductive Sciences (3 vols)*. London: Parker, 1837.
[Whewell, 1840] W. Whewell. *The Philosophy of the Inductive Sciences (2 vols)*. London: Parker, 1840.
[Wright, [1931]1986] S. Wright. Evolution in Mendelian populations. Genetics, 16:2 (1931). In *Evolution: Selected Papers.* Edited by W B Provine. Chicago: Chicago University Press, [1931]1986.
[Wright, [1932]1986] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution . In *Evolution: Selected Papers.* Edited by W B Provine. Chicago: Chicago University Press, [1932]1986.

This page intentionally left blank

SIR RONALD AYLMER FISHER

Robert A. Skipper, Jr.

Sir Ronald Aylmer Fisher (1890–1962) was, along with J. B. S. Haldane (1892–1964), and Sewall Wright (1889–1988), an architect of mathematical population genetics. His contributions to statistics were no less monumental. Indeed, Fisher's work in both areas of science continues its dominance. This biographical essay focuses on Fisher's contributions to evolutionary theory, although it very briefly chronicles his foundational work in statistics. A detailed account of Fisher's life and work may be found in the biography written by one of his six daughters, Joan Fisher Box, entitled *R. A. Fisher: The Life of a Scientist* [1978].

1 EARLY LIFE AND EDUCATION

Fisher was born February 17, 1890 in London, England. Fisher and his twin brother, who died in infancy, were the youngest of eight children born to Katie Heath and George Fisher. R. A. showed his mathematical gifts very early. But because of his extreme myopia, his early training was unconventional. In fact, some of his most influential early tutoring was done without pencil or paper, which would give Fisher a remarkable ability to do complex mathematics in his head. At the Harrow School, for instance, Fisher, at the age of 16, won the Neeld Medal in a mathematical essay competition open to the entire school.

Fisher matriculated at Gonville and Caius College, Cambridge, in 1909. Fisher's interests were broad, including astronomy, mathematics, physics, and biology. In 1911, he worked to found the Cambridge University Eugenics Society, an effort that profoundly influenced his interest in evolutionary biology. Fisher's active interest in evolutionary problems was also influenced by his discovery of Karl Pearson's series of papers collected between 1893 and 1912 as *Mathematical Contributions to the Theory of Evolution.* Fisher graduated from Cambridge in 1912 with a first in mathematics, and would stay on another year to study physics and statistical mechanics. In 1914 he became a statistician at the Mercantile and General Investment Company in London. Contrary to his desires, Fisher was kept out of World War I because of his poor vision. Between 1915 and 1919 Fisher taught mathematics and physics at a variety of public schools. And in 1917 he married Ruth Eileen Gatton Guinness with whom he would parent eight children, two sons and six daughters.

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

2 AT ROTHAMSTED: A PERIOD OF INTENSE ACTIVITY

In 1919, Fisher accepted a statistician's post at Rothamsted Experimental Station, one of the oldest agricultural research institutions in England. There, between 1919 and 1933, Fisher made substantial contributions to statistics and genetics. And by 1929, he was elected to the Royal Society. In statistics, Fisher introduced the concept of "likelihood" in 1921. The likelihood of a parameter is proportional to the probability of the data and it gives a function which usually has a single maximum value, i.e., the maximum likelihood. The next year Fisher introduced a new conception of statistics the aim of which was the reduction of data and the problematic of which was the specification of the kind of population from which the data came, estimation, and distribution. Soon thereafter, in 1925, Fisher published *Statistical Methods for Research Workers*, in which he articulated methods for the design and evaluation of experiments. The 1925 edition was the first of many.

In genetics, Fisher published the first paper, in 1922, adumbrating a mathematical synthesis of Darwinian natural selection with the recently rediscovered laws of Mendelian heredity. In that paper, entitled "On the Dominance Ratio," Fisher discusses, as he says, "the distribution of the frequency ratio of the allelomorphs of dimorphic factors, and the conditions under which the variance of the population may be maintained" [1922, p. 322]. He sees this paper as following on the heels of his 1918 paper, "The Correlation Between Relatives on the Supposition of Mendelian Inheritance." In broad brush strokes, what I take this to mean is that where the 1918 paper defended the principles of Mendelian heredity against the criticisms of the biometricians (and in fact showed the two schemes to be compatible), the 1922 paper continues by carrying through its mathematical methods and concepts as well as defending Darwinism *using* the principles of Mendelian heredity. Specific to "On the Dominance Ratio," Fisher's aim was to respond to a set of criticisms that Darwinian natural selection cannot be the correct explanation of the modulation of genetic variation in populations because the genetics of populations are such that there is not enough variation available for selection to act on. In his response, Fisher considered the interaction of natural selection, random survival (genetic drift), assortative mating, and dominance. During the course of the paper, Fisher eliminated from consideration what he took to be insignificant evolutionary factors, such as epistatic gene interaction and genetic drift, and argued that natural selection acted very slowly on mutations of small effect and in the context of large populations (10^4)maintaining a large amount of genetic variation.

Consider drift, or what Fisher referred to variously as random survival, steady decay, or the "Hagedoorn effect." The phrase "random drift" comes from Wright's 1931 landmark paper, "Evolution in Mendelian Populations." Notwithstanding Wright's obvious contributions to the development of the concept and mathematical modeling of drift, it was Fisher who, in this 1922 paper, was the first among the architects of population genetics to mathematically explore the evolutionary consequences of drift in a Mendelian population.

In finite populations, the variation in the number of offspring between individuals may result in random fluctuations in allele frequencies. These random fluctuations affect the chances of survival of a mutant allele in a population. Fisher argued that the survival of a rare mutant depended upon chance and not selection. Indeed, he argued that such a mutation would be more likely to become fixed at low frequencies in a large rather than in a small population since in a large population the mutant would have a greater probability of survival. Random fluctuations in allele frequencies also reduce a population's genetic variation. In 1921, A. L. and A. C. Hagedoorn argued that random survival is an important cause of the reduction of genetic variation in natural populations. Fisher argued that the Hagedoorns were mistaken. Fisher determined two key quantities for the situation in which a population is under the influence only of the steady decay of genetic variation, that is, the Hagedoorn effect: The first quantity describes the time course in generations of the Hagedoorn effect; the second describes the "half-life" in generations of the effect. Fisher determined the time course to be the product of four and the population size and the "half-life" to be the product of 2.8 and the population size [Fisher, 1922, p. 330]. This means that the Hagedoorn effect requires, in generations, the product of four and the population size to reduce the genetic variation in the population to the point that all alleles are identical by descent. The "half-way" point is reached in the product of 2.8 and the population size generations. (Wright demonstrated in a 1929 letter to Fisher that his, Fisher's, calculations were twice too high: the time-course in generations is the product of 2 and the population size and the "half-life" of the Hagedoorn effect is the product of 1.4 and the population size. In a 1930 paper, Fisher [1930a] showed that the correction had only a minor effect on his argument.)

Fisher used these quantities to "weight" the significance of the effect of steady decay; the longer the time course, the weaker the effect. Given that the time course of the Hagedoorn effect depends on the population size, the larger the population, the weaker, or less significant the effect. It is evident that as population size increases over 10^4, that the time course becomes considerable. Indeed, Fisher says,

> [a]s few groups contain less than 10,000 individuals between whom interbreeding takes place, the period required for the action of the Hagedoorn effect, in the entire absence of mutation, is immense. [1922, p. 330]

According to Fisher, then, the Hagedoorn effect is evolutionarily insignificant and populations are large.

Fisher's insights regarding the evolutionary effects of genetic drift are what is behind his reflection of Darwin when he (Fisher) says,

> a numerous species, with the same frequency of mutation, will maintain a higher variability than will a less numerous species: in connection with this fact we cannot fail to remember the dictum of Charles Darwin,

> that 'wide-ranging, much diffused and common species vary most'. [1922, p. 324]

Consider next gene interaction.

In his 1918 paper, Fisher considered the statistical consequences of dominance, epistatic gene interaction, assortative mating, multiple alleles, and linkage on the correlations between relatives. Fisher argued that the effects of dominance and gene interaction would confuse the actual genetic similarity between relatives. He also knew that the environment could confuse such similarity. Fisher here introduced the statistical concept of variance and the analysis of variance to the scientific literature. He says,

> When there are two independent causes of variability capable of producing in an otherwise uniform population distributions with standard deviations σ_1 and σ_2, it is found that the distribution, when both causes act together, has a standard deviation $\sqrt{\sigma_1 + \sigma_2}$. It is therefore desirable in analyzing the causes of variability to deal with the square of the standard deviation as the measure of variability. We shall term this quantity the Variance of the normal population to which it refers, and we may now ascribe to the constituent causes fractions or percentages of the total variance which they together produce. [1918, p. 399]

Fisher then used this tool to partition the total variance into its component parts.

Fisher labeled that portion of the total variance that accurately described the correlation between relatives the "additive" genetic component of variance. The "non-additive" genetic component included dominance, gene interaction, and linkage. Environmental effects, such as random changes in environment, comprised a third component of the total variance. In 1922, on the basis of the 1918 work, Fisher argued that the additive component of variance was important for evolution by natural selection. Indeed, he argued that, particularly in large populations ($> 10^4$), non-additive and environmental components of the total variance are negligible. He further claimed that selection would remove any factor for which the additive contribution to the total genetic variance is very high and to leave those for which the contribution is low. Indeed, Fisher says,

> It is therefore to be expected that the large and easily recognized factors in natural organisms will be of little adaptive importance, and that the factors affecting important adaptations will be individually of very slight effect. [1922, p. 334]

Ultimately, for Fisher, evolution proceeds very slowly, with low levels of selection acting on mutations of small effect and in large populations (10^4) holding considerable genetic variation.

Fisher's work discussed above and other work on, e.g., the evolution of dominance and mimicry, would culminate in his 1930 book, *The Genetical Theory of Natural Selection*, one of the principal texts, along with those of Haldane [1932] and Wright [1931; 1932], completing the reconciliation of Darwinism and Mendelism and establishing the field of theoretical population genetics (and, for Fisher, its application to eugenics). *The Genetical Theory* is celebrated as the *locus classicus* for the reconciliation. Remarkably, the book manuscript was produced by Fisher dictating to his wife, Ruth, during the evenings. The book was revised and reissued in 1958 and most recently in a variorum edition issued in 1999.

The first seven (of twelve) chapters of *The Genetical Theory* set out Fisher's synthesis of Darwin's mechanism of natural selection and Mendelian genetics. Fisher considered the first two chapters, on the nature of inheritance and the "fundamental theorem of natural selection," the most important of the book. Indeed, these two chapters accomplish the key piece of the reconciliation. Moreover, the general argument strategy Fisher used in 1918 and 1922 of defending the principles of Mendelian heredity and defending Darwinism under the rubric of Mendelian heredity, is carried through. Fisher's aim in *The Genetical Theory* is to establish particulate inheritance against the blending theory and then demonstrate how plausibly Darwinian natural selection may be the principal cause of evolution in Mendelian populations.

Fisher's first chapter considers implications of a synthesis of natural selection with, alternatively, blending and Mendelian inheritance. He demonstrates that on the Mendelian theory, natural selection may be the main cause of a population's variability. The demonstration importantly resolved a persistent problem for Darwin's theory of descent with modification, one that had led biologists to abandon natural selection as an evolutionary cause: Darwin's acceptance of blending inheritance required him to imagine special causes controlling mutation because of enormous mutation rates demanded by the blending theory. Because Mendelian heredity did not demand such enormous mutation rates, Fisher was able to eliminate these controlling causes and, so, revive natural selection as an important evolutionary cause.

Fisher's second chapter develops, mathematically, his genetical theory of natural selection. The arguments are drawn largely from 1922's, "On the Dominance Ratio," and 1930's, "The Distribution of Gene Ratios for Rare Mutations," the response to Wright's aforementioned correction of Fisher's 1922 paper. Three key elements may be distilled from Fisher's "heavy" mathematics in the second chapter of *The Genetical Theory*. The first is a measure of average population fitness, Fisher's "Malthusian parameter," i.e., the reproductive value of all genotypes at all stages of their life histories. The second is a measure of variation in fitness, which Fisher partitions into genetic and environmental components (based on his distinctions from 1918 and 1922). The third is a measure of the rate of increase in fitness, i.e., the change in fitness due to natural selection. For Fisher, "*the rate of increase of fitness of any organism at any time is equal to its genetic variance in fitness at that time*" ([1930b, p. 37], emphasis in original). This last element is

Fisher's "fundamental theorem of natural selection," and is the centerpiece of his natural selection theory.

Interestingly, inasmuch as Fisher considered his fundamental theorem the centerpiece of his evolutionary theory, it happens that the theorem is also the most obscure element of it. The theorem was thoroughly misunderstood until 1989 when Warren Ewens rediscovered George Price's 1972 clarification and proof of it. Fisher's original statement of the theorem in 1930 confusingly suggests that mean fitness can never decrease because variances cannot be negative. Price showed that in fact the left-hand-side of the equation that describes the theorem is not the total rate of change in fitness but rather only one component of it. That part is the portion of the rate of increase that can be ascribed to changes in gene frequencies. And, actually, in Fisher's ensuing discussion of the theorem, he makes this clear. The total rate of change in mean fitness is due to a variety of forces including gene frequencies themselves, environmental changes, epistatic gene interaction, dominance and so forth. The theorem isolates the changes in gene frequencies from the rest, a move suggested in Fisher's 1922 paper. The key change Price and Ewens make in the statement of the theorem, the change that clarifies it, is to write "*additive* genetic variance" for "genetic variance." With the theorem clarified and proven, Price and later Ewens argue that it is not so fundamental. Given that it is a statement about only a portion of the rate of increase in fitness, it is incomplete.

Under the rubric of the fundamental theorem, Fisher offers a geometrical proof of his conclusion from 1922 that cumulative evolution is primarily the result of low pressures of natural selection on mutations of small effect. Fisher's model depicts an organism as a point in a space of very high dimension. Each dimension is a phenotypic trait. For Fisher, adaptation is a step-wise process to the optimum phenotype, and each step is the substitution of an advantageous mutation having some phenotypic effect. Consider Figure 1, which depicts Fisher's model in two dimensions. Two continuous phenotypes, X and Y, have a fitness optimum at O. Fitness decreases the further X and Y move away from the optimum; the circle through D and centered around O represents the phenotypes that have the same low fitness. If a population is at D, mutations of small and large effect arising in the population are represented, respectively, by the small and large circles centered around D. The dashed portion of the circles corresponds to mutations that bring the population from D to points closer to O and are therefore advantageous. The solid portion corresponds therefore to deleterious mutations, which take the population away from O. In spite of the fact that the advantageous proportion is about half of the small circle, it is much less than half of the large one. Consequently, evolution by natural selection is more likely to proceed by small steps because advantageous mutations of small effect will be more common than advantageous mutations of large effect. And the higher the dimensionality of the model, the greater the bias against mutations of large effect. (Discussion follows [Burch and Chao, 1999].)

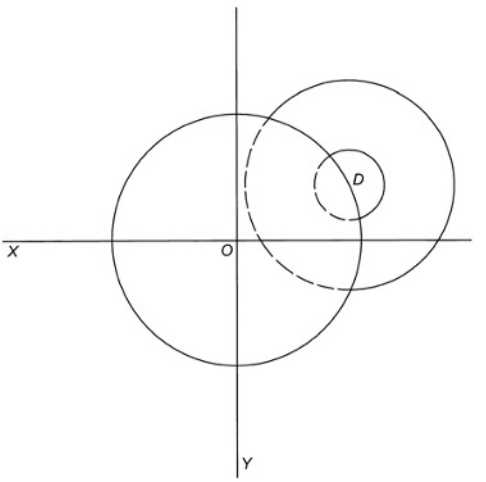


Figure 1. Graphical Representation of Fisher's Geometric Model. Adapted from Burch and Chao [1999, 922]. X and Y are continuous phenotypes with optimum fitness at O. D is a non-optimal point of fitness. See text for discussion. Fisher himself does not use a diagram although he does articulate, mathematically, the relevant properties.

By and large, the middle chapters of *The Genetical Theory* are (mainly theoretical) explorations of cases, such as dominance, sexual selection, and mimicry, to support and extend the preceding theoretical work. Nevertheless, in the fourth and fifth chapters, Fisher expands his theoretical discussion to more general issues concerning the causes of genetic variation. The last five chapters of the book explore natural selection in human populations, particularly social selection in human fertility. Fisher's central observation, based upon England's 1911 census data, was that the development of economies in human societies structures the birth-rate so that it is inverted with respect to social class. In the final chapter of his book, Fisher offers strategies for countering this effect. He proposed the abolishment of the economic advantage of small families by instituting what he called "allowances" to families with larger numbers of children, with the allowances proportional to the earnings of the father. In spite of Fisher's espousal of eugenics in this part of the book, he means the discussion to be taken as an inseparable extension of the preceding part.

Fisher compared both his 1922 and 1930 exploration of the balance of evolutionary factors and the "laws" that describe them to the theory of gases and the second law of thermodynamics, respectively. Of the 1922 investigation, Fisher says,

> [t]he investigation of natural selection may be compared to the analytic treatment of the Theory of Gases, in which it is possible to make the

> most varied assumptions as to the accidental circumstances, and even the essential nature of the individual molecules, and yet to develop the natural laws as to the behavior of gases, leaving but a few fundamental constants to be determined by experiment. [1922, pp. 321-322]

He continues the analogy in 1930, adding that

> the fundamental theorem... bears some remarkable resemblances to the second law of thermodynamics. Both are properties of populations, or aggregates, true irrespective of the nature of the units which compose them; both are statistical laws; each requires the constant increase in a measurable quantity, in the one case the entropy of the physical system and in the other the fitness... of a biological population... Professor Eddington has recently remarked that 'The law that entropy always increases — the second law of thermodynamics — holds, I think, the supreme position among the laws of nature'. It is not a little instructive that so similar a law should hold the supreme position among the biological sciences. [1930b, pp. 36-37]

The received view of these comparisons is that Fisher's interests in physics and mathematics led him to look for biological analogs (e.g., Provine 1971, Gayon 1998). No doubt this is part of the story. However, I think a more plausible interpretation of the comparison comes from treating Fisher's major 1918, 1922, and 1930 works as one long argument. If we do so, as I have roughly done in the preceding, we find that Fisher's strategy in synthesizing Darwinian natural selection with the principles of Mendelian heredity was to defend, against its critics, selection as an evolutionary cause under Mendelian principles. Following this argument strategy, Fisher built his genetical theory of natural selection piece meal, or from the bottom up. That is, Fisher worked to justify the claim of his fundamental theorem by constructing plausible arguments about the precise balance of evolutionary factors. Thus, his piece meal consideration of the interaction between dominance, gene interaction, genetic drift, mutation, selection, etc. led to his theorem. It was not, at least not primarily, the search for biological analogues to physical models and laws that underwrites the theorem.

The Genetical Theory of Natural Selection is a point of departure in contemporaneous evolutionary thought, responsible in part for the origination of theoretical population genetics and what is commonly called the "modern synthetic theory of evolution." *The Genetical Theory* was followed by Wright's and Haldane's major such works in 1931 and 1932. It is useful to consider, briefly, Fisher's contribution against the background of Haldane's and Wright's.

Haldane's 1932 book, *The Causes of Evolution*, is the capstone text not only of Haldane's foundational series of 10 mathematical evolution papers he published between 1924 and 1934, it was the capstone text of the origins of theoretical population genetics. Haldane's problematic was aligned with his contemporaries; he wanted to understand the nature and significance of natural selection in Mendelian

populations. Haldane is closer to Fisher than he is to Wright. But Haldane's support of Fisher is not in the form of an uncritical summary. Haldane carried the mathematical exploration of selection in finite population sizes, changing environments, and in multiple dimensions further than Fisher had, and in part as a critical consideration of Wright's work on the topic. Moreover, and importantly, Haldane emphasized the problem of the units of selection, recognizing that selective forces acting on gametes, organisms, or populations may come into conflict. Today, the book is known primarily for its critical comparison of Fisher's, Wright's, and Haldane's own mathematical explorations of this issue. But the book is much more than that. Haldane accomplishes a synthesis not merely of Darwinism and Mendelian heredity, but also, under the rubric of population genetics, of chromosomal mechanics, cytology, and biochemistry. (See, e.g., [Sarkar, 2004].)

Wright's 1931 paper, "Evolution in Mendelian Populations," is his main early paper demonstrating the mathematical unification of Darwinian natural selection and the principles of Mendelian heredity. Wright communicated this synthesis in the form of his Shifting Balance Theory of evolution. Wright would use his famous "adaptive landscape" diagram to communicate his theory a year later, in 1932. Flowing from his work as a physiological geneticist, Wright emphasized genetic interaction in evolution by way of the following assumption: Because the field of gene combinations in a population are vast, genes adaptive in one combination may not be adaptive in another; consequently, the field of joint gene frequencies graded for adaptive value will produce a "hilly landscape." Populations thus face the problem of shifting from one peak to another. Wright solved the problem of "peak shifts" with his three-phase shifting balance process. In the first phase, subdivided populations would be subject to genetic drift, pulling them into a "valley." In the second phase, those populations would be dragged up the next peak by intrademe selection, or within-group selection. And in the third phase, the global population would find the highest peak by interdeme selection, or selection driven by differential dispersal.

Wright's mathematical analysis of the reconciliation between Darwinism and Mendelism agreed with Fisher's and Haldane's. However, Fisher and Wright disagreed about the extent to which epistatic gene interaction, genetic drift, and population structure were important in accounting for cumulative evolutionary change. Where Wright imagined a "hilly" adaptive landscape, requiring a constellation of evolutionary factors to traverse it, Fisher saw a landscape with a single peak, requiring only selection and mutation to traverse it. And beginning in 1929, Wright and Fisher became mired in controversy over their alternative understandings of the evolutionary process that would last until Fisher's death in 1962 [Provine, 1986]. But Fisher's death did not end the controversy. Fisher's and Wright's acolytes, as well as Wright himself (until his death in 1988), would continue exploring the differences between the apparently alternative evolutionary theories. Between 1997 and 2000, in fact, a dispute erupted reminiscent of the debates between Fisher and Wright between the team of Jerry Coyne, Nicholas Barton, and Michael Turelli [1997; 2000] and the team of Michael Wade and Charles

Goodnight [1998; 2000]. Coyne *et al.*. develop a systematic critique of Wright's Shifting Balance Theory, arguing that it is theoretically problematic and that it lacks empirical support. They conclude that the Fisherian perspective is preferable. Wade and Goodnight counter some of Coyne et al.'s criticisms and argue that there is ample room in the evolutionary domain for both the Fisherian and Wrightian perspectives. ([Skipper, 2002] provides an analysis of the debate).

3 AT UNIVERSITY COLLEGE, LONDON

Fisher would leave Rothamsted to take the Galton Chair of Eugenics at University College, London in 1933. Karl Pearson, retiring the Chair, appointed Fisher and Pearson's son, Egon, to share the post. Fisher controlled the genetics section; Pearson controlled the statistics section. Fisher had turned down Pearson's offer of chief statistician at Galton Laboratories for the position at Rothamsted in 1919. And by the time he took the Galton Professorship at University College, he was thoroughly mired in controversy over the foundations of statistics with Karl Pearson and his followers. Fisher's revolutionary work in statistics came at around the time that Pearson's own work was showing weaknesses; unfriendly competition catalyzed the controversy, and Pearson would take ill-feelings toward Fisher, who had his own ill-feelings toward Pearson, to his grave.

Until 1934, Fisher remained active with the Eugenics Society. Indeed, Fisher was involved in the society's campaign, starting in 1929, for a law allowing voluntary sterilization on the basis of the claims of eugenics Fisher and others, such as John Maynard Keynes, R. C. Punnett, and Charles Darwin's son Horace Darwin, advocated. The draft legislation was soundly defeated in Parliament. Fisher resigned from the Society in 1934, along with other members, due to the group's infighting over the ever-decreasing role of its scientific members.

In 1935, Fisher's *The Design of Experiments* appeared and, like *Statistical Methods for Research Workers*, would be expanded and reissued many times. These two works, and Fisher's (with Frank Yates) 1947 *Statistical Tables for Biological, Agricultural, and Medical Research* revolutionized agricultural research. Also in 1947, Fisher would found, with the geneticist Cyril Darlington, the journal *Heredity*. Its first volume would include a publication resulting from a collaboration Fisher entered into with the ecological geneticist E. B. Ford on the evolution of the Scarlet Tiger moth, *Panaxia dominula*. This publication, "The Spread of a Gene in Natural Conditions in a Colony of the Moth, *Panaxia dominula*," bears mentioning for several reasons. The first is that the paper describes a field study aimed at discrediting Wright's view that random genetic drift is an important evolutionary factor. As such, it is part and parcel of the controversy between the two. Second, the paper articulates one of the earliest uses of the capture-release protocol for exploring changes in gene frequency in a population. Third, the paper is a cornerstone of the Oxford School of Ecological Genetics founded by Ford. And finally, the 1947 paper is important is because it is the longest-running purported demonstration, at six decades, of natural selection occurring in the wild.

4 AT LIFE'S CLOSE

To be sure, Fisher's work in statistics was revolutionary at the field's conceptual foundations. Moreover, Fisher's work in genetics, highlighted mainly by his 1930 *The Genetical Theory of Natural Selection*, would, with good company in Haldane and Wright, revolutionize biology. Fisher's accomplishments did not go unrecognized. In 1938, he was awarded the medal of the Royal Society, in 1948, the Darwin medal, and in 1955 the Copley medal. In 1943 Fisher was appointed Balfour Professor of Genetics at Cambridge University upon the retirement of the geneticist R. C. Punnett. The department had been decimated due to World War II. However, Fisher was able to recruit, in 1948, Luigi Luca Cavalli-Sforza, the famed Italian population geneticist known especially for his investigations of human genetic diversity. In 1952 Fisher was dubbed Knight Bachelor by Queen Elizabeth II. And in 1957 Fisher became President of his alma mater, Gonville and Caius College, Cambridge. These are only some of the honors bestowed upon Fisher.

After a luminous career as statistician and biologist, Fisher died at the age of 73 on July 29, 1962 in Adelaide, Australia where he had retired in 1957 to take a post as a senior research fellow at the Commonwealth Scientific and Industrial Research Organization. Fisher the man has been described as charming and warm among friends, but with a volatile temper. Fisher's temper, combined with an unwavering commitment to his views, drove him to heated controversies with other scientists. Fisher's writings have been described as difficult, with much of what he contributed more effectively conveyed by others who were able to simplify Fisher's presentation, but at the same time as genius.

BIBLIOGRAPHY

[Burch and Chao, 1999] C. Burch and L. Chao. Evolution by Small Steps and Rugged Landscapes in the RNA Virus ϕ6, *Genetics* 151: 921-927, 1999.

[Coyne *et al.*, 1997] J. A. Coyne, N. H. Barton, and M. Turelli. Perspective: A Critique of Sewall Wright's Shifting Balance Theory of Evolution, *Evolution* 51: 643-671, 1997.

[Coyne *et al.*, 2000] J. A. Coyne, N. H. Barton, and M. Turelli. Is Wright's Shifting Balance Process Important in Evolution? *Evolution* 54: 306-317, 2000.

[Ewens, 1989] W. Ewens. An Interpretation and Proof of the Fundamental Theorem of Natural Selection, *Theoretical Population Biology* 36: 167-180, 1989.

[Fisher Box, 1978] J. Fisher Box. *R. A. Fisher: The Life of a Scientist*. London, UK: Wiley & Sons, 1978.

[Fisher, 1921] R. A. Fisher. On the 'Probable Error' of a Coefficient of Correlation Deduced from a Small Sample, *Merton* 1: 3-32, 1921.

[Fisher, 1922] R. A. Fisher. On the Dominance Ratio, *Proceedings of the Royal Society of Edinburgh* 42: 321-341, 1922.

[Fisher, 1925] R. A. Fisher. *Statistical Methods for Research Workers*. London: Oliver & Boyd, 1925.

[Fisher, 1930a] R. A. Fisher. The Distribution of Gene Ratios for Rare Mutations, *Proceedings of the Royal Society of Edinburgh* 50: 205-220, 1930.

[Fisher, 1930b] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford University Press, 1930. Released in a revised, second edition in 1958 from Dover Publications. Released in a variorum edition by Oxford University Press in 1999 edited by J. H. Bennett.

[Fisher, 1935] R. A. Fisher. *The Design of Experiments*. London: Oliver & Boyd, 1935.
[Fisher and Ford, 1947] R. A. Fisher and E. B. Ford. The Spread of a Gene in Natural Conditions in a Colony of the Moth, *Panaxia dominula*, *Heredity* 1: 143-174, 1947.
[Fisher and Yates, 1947] R. A. Fisher and F. Yates. *Statistical Tables for Biological, Agricultural, and Medical Research*. London: Oliver & Boyd, 1947.
[Gayon, 1998] J. Gayon. *Darwinism's Struggle for Survival: Heredity and the Hypothesis of Natural Selection*. Cambridge, UK: Cambridge University Press, 1998.
[Goodnight and Wade, 2000] C. J. Goodnight and M. J. Wade. The Ongoing Synthesis: A Reply to Coyne, Barton, and Turelli, *Evolution* 54: 317-324, 2000.
[Hagedoorn and Hagedoorn, 1921] A. L. Hagedoorn and A. C. Hagedoorn. *The Relative Value of the Processes Causing Evolution*. M. Nijoff: The Netherlands, 1921.
[Haldane, 1932] J. B. S. Haldane. *The Causes of Evolution*. London: Longmans, 1932.
[Price, 1972] G. R. Price. Fisher's 'Fundamental Theorem' Made Clear, *Annals of Human Genetics* 36: 129-140, 1972.
[Provine, 1971] W. Provine. *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press, 1971.
[Provine, 1986] W. Provine. *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press, 1986.
[Sarkar, 2004] S. Sarkar. Evolutionary Theory in the 1920s: The Nature of the Synthesis, *Philosophy of Science* 71: 1215-1226, 2004.
[Skipper, 2002] R. A. Skipper, Jr. The Persistence of the R. A. Fisher-Sewall Wright Controversy, *Biology and Philosophy* 17: 341-367, 2002.
[Wade and Goodnight, 1998] M. J. Wade and C. J. Goodnight. Perspective: The Theories of Fisher and Wright in the Context of Metapopulations: When Nature Does Many Small Experiments, *Evolution* 52: 1537-1553, 1998.
[Wright, 1931] S. Wright. Evolution in Mendelian Populations, *Genetics* 16: 7-159, 1931.
[Wright, 1932] S. Wright. The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution, *Proceedings of the Sixth Annual Congress of Genetics*: 356-366, 1932.

HALDANE AND THE EMERGENCE OF MODERN EVOLUTIONARY THEORY

Sahotra Sarkar

1 INTRODUCTION

This paper is about Haldane's place in the development of evolutionary theory, an attempt to assess his role in the so-called modern synthesis, but it also delves into the more philosophical issue whether Waddington's and Mayr's well-known critique of mathematical population genetics is credible.[1] It looks, in some detail, at what Haldane achieved in the 1920s and early 1930s but ignores important later contributions, in particular, his introduction of genetic load arguments.[2] The paper is partly motivated by the fact that the two standard historical accounts of the development of modern evolutionary theory pay little attention to Haldane. The first, the "genetical" account, most systematically elaborated by Will Provine [1971; 1985; 1986], sees the history as an epic struggle between Fisher and Wright and their respective followers. In sharp contrast, the second "naturalist" account, advocated since the late 1950s by Ernst Mayr, has both a negative and a positive thesis. The negative thesis is the rejection of any important role for theoretical population genetics. The positive thesis is an alternative reconstruction of the history of twentieth-century evolutionary theory.

For instance, Mayr [1980, 1] inaugurated a conference on the history of the so-called synthesis with the following remarks:

> The term 'evolutionary synthesis' was introduced by Julian Huxley in *Evolution: The Modern Synthesis* (1942) to designate the general acceptance of two conclusions: gradual evolution can be explained in terms of small genetic changes ('mutation') and recombination, and the ordering of this genetic variation by recombination; and the observed evolutionary phenomena, particularly macroevolutionary processes and speciation, can be explained in a manner that is consistent with the known genetic mechanisms. The objective of this conference is to examine the rapid changes in evolutionary biology that occurred in the period of the synthesis (from approximately 1936 to 1947).

[1] Sarkar [2005] takes up the issue whether the development of modern evolutionary theory should be viewed as a synthesis — that question is ignored here.

[2] See Haldane [1937; 1957]. Crow [1992] has particularly emphasized the importance of these contributions.

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

The passage contains two errors: Huxley [1942] never used the term "evolutionary synthesis" and the concept of such a synthesis goes back to Haldane [1938]. But what is crucial is that Mayr frames the synthesis between 1936 to 1947. The definitive works of the theoretical population geneticists were all published by 1932 and Mayr was intentionally shifting attention away from them to figures such as Dobzhansky [1937], Huxley [1940] and, especially, himself [Mayr, 1942]. The naturalists, including the systematicists, emerge as the dominant figures in his account.

Thus Mayr, too, has no place for Haldane. This is particularly striking because some commentators, including Carson [1990, 89] at the same conference, argued that Haldane's [1932] *Causes of Evolution* was central to the synthesis. As Carson put it, that book was the only one in that era in which

> we find integrated and facile discussions of the evolutionary implication of the *Drosophila* chromosome studies, allopolyploidy in *Primula*, and Darwinian fitness[.] Here Haldane neatly conjoins Darwin and Mendel, Fisher and Wright, Newton and Kihara. In the evolutionary context, Haldane deals for the first time with inversions and translocations, polyploidy and hybridization. The paleontological record is woven into the argument. [1990, 89].

Moreover, in a well-known mathematical appendix of that book which served as a primer for the next generation of population geneticists, Haldane collected and systematized all the results that he, Fisher, and Wright had thus far obtained. It is time for a reappraisal of Haldane's role.

2 HISTORICAL BACKGROUND

To see what Haldane's work achieved in the 1920s and 1930s requires some discussion of the background. Mendel's laws, which had remained unknown for a generation, had been recovered only around 1900. A systematic exploration of their mathematical consequences began immediately with G. Udny Yule [1902] and, especially, Karl Pearson [1904] who, however, was a strident advocate of the competing biometrical model of inheritance. Pearson was unimpressed, arguing that Mendel's laws gave values far too low for the correlation coefficients between traits in related individuals which his biometrical laboratory had been measuring for the past decade. But these arguments were far from conclusive with Yule [1906] pointing out that a straightforward relaxation of one of Pearson's assumptions (complete dominance) resolved the apparent discrepancies between the observed and calculated coefficients.

While this dispute stagnated in the background, scattered work on a variety of Mendelian models began, primarily in Britain and the United States. The first truly significant result of theoretical population genetics, the Hardy-Weinberg rule, was independently discovered in 1908 by a German physician, W. Weinberg, and

the well-known British mathematician, G. H. Hardy.[3] Suppose that a trait is controlled by one gene (or "locus") which takes one of two forms or "alleles", A or a. Then each diploid organism has either AA, Aa or aa as its genotype. Suppose p and q are the relative frequencies of A and a, respectively, in the population in some generation ($p + q = 1$). Assuming that the population is large (in principle, infinite), that it mates at random with respect to this trait, and that there is no selection, the Hardy-Weinberg rule stated that the genotypes AA, Aa and aa have frequencies p^2, $2pq$ and q^2, respectively, in the next and all subsequent generations. The effect of selection, non-random mating, and other such factors would be to drive a population away from the Hardy-Weinberg ratios.[4] Weinberg [1909a,b; 1910] continued the mathematical exploration of Mendelism and went on to argue — as had Yule before him — that Pearson's negative assessment was unjustified. Pearson [1910] remained unimpressed, accusing Weinberg of giving a "curiously ignorant account of the biometric treatment of heredity" [p. 381n].

Meanwhile, in 1910, Hardy's example immediately led a young mathematician at Cambridge, H. T. J. Norton, to write the first thesis in mathematical genetics. Persistent ill health prevented him from immediately publishing most of his results. However, he calculated a "selection table" for R. C. Punnett's [1915] *Mimicry in Butterflies*. This table showed that even mild selection could be quite effective in promoting the rapid spread of an allele in a population. For those who worried about whether there had been enough time for natural selection to have created the observed patterns of evolutionary change, Norton's table provided reassurance, though only in the very restricted context of what he had modeled.

Meanwhile, in the United States, the mathematical exploration of consequences of Mendelian inheritance began around 1912 with the work of H. S. Jennings [1912] who first calculated the results of self-fertilization, rather than random mating, on the genetic composition of a population, and then generalized this treatment to study inbreeding [Jennings, 1914]. Pearl [1913; 1914a,b] and Fish [1914] continued to explore these models.[5] In 1916, going beyond inbreeding and autosomal factors, Jennings considered the effects of "assortative" or preferential mating between like phenotypes and sex-linked factors [Jennings, 1916]. Wentworth and Remick [1916] extended some of these results.

Jennings [1917] extended his arguments to a pioneering discussion of two locus models. Let A, a and B, b pairs of alleles at two loci. He found out that, if the system of mating was known, it was possible to calculate the frequencies of the

[3]Pearson [1904] had already proved a special case of the same result (when both alleles have the same frequency), but Weinberg [1908] and Hardy [1908] deserve credit for recognizing and stating it in its full generality.

[4]However, as Hardy [1908], noted the ratios were "stable" in the sense that, not only did they remain the same generation after generation in the absence of interfering factors, if there was a deviation in some generation, the distribution would simply shift to another set of frequencies (say, p'^2, $2p'q'$, q'^2) which would also persist ever afterwards.

[5]Pearl [1913] incorrectly suggested that inbreeding other than self-fertilization did not necessarily increase the frequency of homozygotes but subsequently corrected this error and gave formulae for the effects of brother-sister mating [Pearl, 1914a,b]. H. D. Fish, who was the first to notice Pearl's error, went on to explore mating between parent and offspring [Fish, 1914].

16 possible genotypes from the frequencies of the gametes AB, Ab, aB and ab. It sufficed, therefore, only to consider the simpler problem of the change in the gametic frequencies. In two papers, Robbins [1917; 1918a] generalized Jennings's results. In a third, and very important contribution, he turned to two loci and proved two important results [Robbins, 1918b]:

(i) following Robbins's notation, let r be the proportion of gametes produced without recombination compared to those that are produced with it. r is, then, a measure of linkage.[6] Let mating be at random at both loci. Then, if p_n, q_n, s_n and t_n are the frequencies of the four gametic types, AB, Ab, aB and ab, respectively, in the n-th generation, then the quantity, $\Delta_n = q_n s_n - p_n t_n$, can be used to write down particularly simple relations for the change in gametic frequencies:

$$p_n = p_{n-1} + \frac{\Delta_{n-1}}{1+r};$$

$$q_n = q_{n-1} - \frac{\Delta_{n-1}}{1+r};$$

$$s_n = s_{n-1} - \frac{\Delta_{n-1}}{1+r};$$

$$t_n = t_{n-1} + \frac{\Delta_{n-1}}{1+r};$$

(ii) the quantity Δ_n changes by a factor of $r/(1 + r)$ each generation, that is,

$$\Delta_n = \frac{r}{1+r}\Delta_{n-1}.$$

The second result is important because it shows that, under the assumptions of the model, Δ_n gradually goes to 0 as the number of generations increases. However, it is strictly equal to 0, and remains so, only if $\Delta_0 = 0$. It thus provides an analog of the Hardy-Weinberg rule for two loci. Robbins's results did not receive the attention they deserved.[7]

Both Jennings and Robbins published analyses of complete selection (that is, complete elimination of the less fit type) which ignored the possibility of weaker selection, measured by a fitness parameter. However, in unpublished work, for

[6] In spite of Robbins's notation, r should not be confused with the recombination fraction. If ρ is the recombination fraction, $\rho = 1/(1 + r)$. With this substitution, Robbins's results assume the standard form of Haldane [1926]. (For a modern appraisal, see Edwards [1977, 94–98]).

[7] Apparently unaware of Robbins's work, Haldane [1926) rederived the first of these two results in the third part of "A Mathematical Theory of Natural and Artificial Selection". The value of the parameter Δ_n would not be fully appreciated until the 1950's when -Δ_n came to be called "linkage disequilibrium" (or, alternatively, "gametic phase disequilibrium"). Robbins also gave more complicated formulae when the linkage between the loci was not the same in the two sexes, and in the final installment of his series of papers, discussed a model of "disassortative mating" in which brother-sister mating was avoided [Robbins, 1918c].

a one locus model with random mating, Jennings extended his analysis to some simple cases of this type; his results were presented by Warren [1917]. Thus, these analyses of selection remained quite rudimentary though they were sufficient to convince Warren that natural selection acting upon Mendelian factors was quite sufficient to explain the type of adaptation that was the result of evolutionary change.

The rudimentary state of the analysis of selection was due for rapid change with the advent of Fisher, Wright, and Haldane. In 1918, Fisher published a long paper, "The correlation between relatives on the supposition of Mendelian inheritance." In it, he systematically carried out a program that Pearson [1904] had perfunctorily abandoned, namely, that of connecting biometry to Mendelism. Giving a far more general analysis than Yule [1906], he showed that all the correlation coefficients measured by Pearson and his collaborators could be systematically and correctly predicted from a Mendelian basis.[8] He also showed that phenotypic traits which depended on a large number of Mendelian factors, each with a tiny effect, would have a normal distribution. This was precisely one of the generalizations that the biometricians had made from their observations. Finally, Fisher showed that the so-called "Law of Ancestral Heredity", by which the biometricians expressed the contribution of each preceding generation to a particular trait, could also be derived from a variety of Mendelian models. If there still were any lingering doubt that biometry and Mendelism were compatible, Fisher's paper removed them. In effect, Fisher had reduced biometrical theory to Mendelism.[9]

In 1922, Fisher turned to the question whether natural selection and mutation sufficed to explain the observed distributions of alleles (at a single locus) in a population. He focused on the factors that maintained variability, being skeptical of the suggestion by A. L. Hagedoorn and A. C. Hagedoorn-Vorstheuvel la Brand [1921] that random survival, rather than selection or mutation, was the critical factor in allele frequency changes. First, Fisher showed that, in the absence of mutation and ignoring stochastic factors, a stable equilibrium between two alleles is only reached if selection favors the heterozygote over the two homozygotes. This was the first explanation of such polymorphism on the basis of what came to be known as "heterosis". Next, he calculated that, if the frequency of any allele is very small, selection cannot prevent random extinction. He then assumed that the alleles had achieved an equilibrium distribution. He calculated the shape of this distribution if fortuitous extinction or various types of selection were balanced by mutation. A low mutation rate sufficed to counteract extinction and maintain genetic variability. With selection, a large mutation rate was necessary if the variation in the population is to be maintained at the equilibrium levels.

Fisher also showed that strongly selected alleles would be rapidly fixed, leaving as variable only alleles that were subject to weak selection. However, almost all

[8] Many of these results had already been obtained by Weinberg [1909a,b; 1910], but in an age when genetics was almost entirely the domain of the Anglophone world, Weinberg's efforts were ignored.

[9] Sarkar [1998; 2004] discusses this reduction in detail.

alleles were supposed to be only subject to weak selection for two reasons: "(1) the effect of the factor on development may be very slight, or (2) the factor may effect changes of little adaptive importance". Therefore, he concluded, "variation of adaptive importance should be due to numerous factors, which individually are difficult to detect" (p. 24). The assumptions that weak selection was paramount and that evolution was most likely to occur in large populations became the two major tenets of Fisher's theory of evolution.

Meanwhile, Wright had just published his first analysis of selection In 1921, in a set of five papers, "Systems of Mating".[10] Unlike Fisher, Wright worked with explicit genetic models (with two alleles at each of one or two loci). However, he, too, assumed a population at equilibrium where the Hardy-Weinberg rule held and, for the two locus case, $\Delta_0 = 0$. He first worked out several of the correlation coefficients for populations at equilibrium — these were only special cases of Fisher's [1918] more general treatment [Wright, 1921a]. However, he gave a general and systematic treatment of inbreeding [Wright, 1921b]. He analyzed assortative mating for a simple two-locus model [Wright, 1921c] but the analysis of selection remained superficial [Wright, 1921d]. When only one locus was involved, he rederived the results of Jennings [1916] and Wentworth and Remick [1916]. For two loci, his results demonstrated that selection decreased the variability within a population. The most lasting contribution of this set of papers was the first systematic presentation of his method of path coefficients, his novel — and peculiar — method for calculating the correlations between variables.[11]

[10] By this point, significant work on population genetics had also begun in the Soviet Union. S. Chetverikov, in Moscow, published an important theoretical paper in 1926, and the group around him even began investigating the genetics of several natural populations of various Drosophila species that were native to the Moscow region [Adams, 1968; 1980]. In the 1920's this work remained unknown in Britain or the United States, where most of the significant development of population genetics theory took place. Chetverikov's 1926 paper was only available in Russian, and a projected English translation was never published probably because of his mysterious arrest by the Soviet secret police (OGPU) in 1929, and subsequent three-year exile to Sverdlovsk and Voronezh [Adams, 1980]. Among Western geneticists, only Haldane seems to have been aware of the Russian work. He met Chetverikov at the Fifth International Congress of Genetics in Berlin in 1927. Subsequently Haldane visited the Soviet Union, as Vavilov's guest, and there is little doubt that he became familiar with Chetverikov's work. Eventually, Haldane had Chetverikov's papers translated into English by a Mrs. A. Sproule, and these were then circulated in his group [Adams, 1980]. However, that was in the 1930's, after he had moved to University College (London) from Cambridge. In 1932, in *The Causes of Evolution*, the only reference to Chetverikov was to his experimental work. The significance of Chetverikov's theoretical paper remained unrecognized during the critical period in the late 1920's when Fisher, Haldane and Wright were systematizing their alternative views about the process of evolutionary change.

[11] In the early 1920s, Wright developed quantitative models of selection. A long manuscript reporting this work remained unpublished, as Wright dealt with his heavy teaching responsibilities at the University of Chicago [Provine, 1986]. A large portion of it became obsolete as, starting in 1924, Haldane began publishing his series of ten papers, all but one of which were entitled, "A Mathematical Theory of Natural and Artificial Selection". The rest of Wright's manuscript, after significant development, became Wright's classic "Evolution in Mendelian Populations", his alternative to Fisher and, incidentally, to Haldane.

3 HALDANE'S "MATHEMATICAL THEORY"

A possible apocryphal story told by Haldane himself states that his interest in genetics began at the age of 10 when, in 1901, his father, the physiologist, J. S. Haldane, had taken him to hear A. D. Darbishire lecture on the then new Mendelism. In any case, while still in school, Haldane noticed evidence of linkage in Darbishire's data on mice. He carried out breeding experiments with the collaboration of his sister and a friend (A. D. Sprunt) but, due to the outbreak of World War I, these results were only published in 1915. It was thus the first discovery, though not the first published report, of linkage in mammals. During part of the war Haldane was convalesced to New Delhi, where he read the reports on Drosophila genetics from Morgan's laboratory at Columbia University. Extending a method of Trow [1913; 1916], he devised the first explicit mapping function to translate recombination frequencies to distances between loci on chromosomes [Haldane, 1919].[12] In 1922, Haldane also published his rule of interspecific hybrids: "When in the F_1 offspring of two different animal races [species] one sex is absent, rare, or sterile, that sex is the heterozygous [heterogametic] sex" [1922, 101]. But, in spite of his long-standing interest in genetics, until 1923, the mathematical problems of natural selection had escaped his attention.

Why that changed so drastically during that year remains unclear. There is compelling circumstantial evidence that part of the answer lies in religious objections to evolution on the ground that natural selection is insufficient as a mechanism to account for all of the past evolutionary changes. The early 1920s witnessed a spirited public controversy between H. G. Wells and Hilaire Belloc over Darwinism. Belloc's religiosity — he hated Wells' materialism — led to a rejection, not of evolution, but of natural selection. Meanwhile, Bateson's and other geneticists' continued doubts about natural selection, as well as efforts to ban the teaching of evolution in some US states, generated ample public controversy about the status of that theory.[13] The prominent anatomist, Arthur Keith [1922a,b], stepped into the dispute. In the *Rationalist Annual* Keith exhorted fellow "Darwinists" to popularize their views. The "very fact that Mr. Chesterton and Mr. Hilaire Belloc could confidently assure readers of the Sunday Press that Darwin's theory was dead", Keith [1922b] argued, "showed that those who are studying the evidence of our origin, and who are Darwinists to a man, had lost touch with public intelligence". Five years later, Haldane rose to Keith's call and also published a piece in the *Rationalist Annual* defending and explaining Darwinism.[14] Eventually, *The Causes of Evolution* would develop that argument in detail.

[12] Wimsatt [1992] examines this work in its historical context, as well as its philosophical underpinnings.

[13] See Belloc [1920a,b], Bateson [1922], Huxley [1922], Keith [1922a], Livingstone [1922], and Robinson [1922].

[14] See Haldane [1927c]. This interpretation of the history was originally put forward by McOuat and Winsor [1995]. The extent to which natural selection had fallen into disrepute was emphasized by Bowler [1989] who argues that the evolutionary synthesis should be regarded as a Mendelian rather than Darwinian revolution.

It is possible that Haldane's mathematical exploration of natural selection, starting 1924, was also a response to Keith's appeal. However, there is also a more direct possible source for Haldane's new-found interest in the mathematics of selection. Among his fellow residents at Trinity College at Cambridge (where Haldane had become Reader in Biochemistry) was Norton. From the latter Haldane learnt not only of Norton's unpublished work, but also surmised that, the efforts of Jennings, Robbins and Warren notwithstanding, the mathematical theory of natural selection remained rudimentary. At this point, Haldane was unaware of either Fisher's or Wright's work. Moreover, neither Fisher nor Wright had yet developed detailed models of selection for specific circumstances that were potentially subject to experimental test. At Cambridge, thanks to Punnett and Norton, that became the focus of Haldane's efforts.

In Haldane's work, Punnett's concern with the time taken by selection to establish a Mendelian variant translated into Haldane's consistent preoccupation with rates of change. Haldane does not start with an assumption of equilibrium and then worry about deviations from it. His attitude even to the Hardy-Weinberg rule was almost dismissive [1924a, 22]: it is merely a way for calculating genotypic ratios in the absence of selection. Unlike Fisher, there was no commitment either to the maintenance of variation in populations or to small selective differences. Unlike what Wright would eventually emphasize, there was no commitment either to a particular breeding or population structure. Initial conditions, whether they be allele frequencies, selective differences, breeding structure, or mode of inheritance, that is, whether the organisms were haploid, diploid (that is, strictly Mendelian) or polyploid, could all be arbitrary. From Haldane's catholic point of view, all options were up for exploration.

The framework for Haldane's [1924a] theory of selection was clearly articulated in the first part of a ten-paper series [1924, 19]:

> A SATISFACTORY theory of natural selection must be quantitative. In order to establish the view that natural selection is capable of accounting for the known facts of evolution we must show not only that it can cause a species to change, but that it can cause it to change at a rate which will account for present and past transmutations. In any given case we must specify:
>
> 1. The mode of inheritance of the character considered,
> 2. The system of breeding in the group of organisms studied,
> 3. The intensity of selection,
> 4. Its incidence (e. g., on both sexes or only one), and
> 5. The rate at which the proportion of organisms showing the character increases of diminishes.
>
> It should then be possible to obtain an equation connecting (3) and (5).

This was the first explicit recipe for the construction of models of natural selection. Following this recipe, Haldane constructed thirteen sets of such models in the first paper and about thirty more in the next nine. The models cover a bewildering variety of situations. Three themes unify the treatment: (i) Haldane insists on explicitly quantitative results; (ii) much more than any of his peers, he attempts to connect each model with experimental data; and (iii) the question of the speed at which selection can produce change always lurks in the background. The discussion that follows will emphasize these themes while navigating through the bewildering variety mentioned above.

What is initially surprising about Haldane's characterization of evolutionary theory above is that he did not distinguish between the relative roles of the selection intensity and the rate of change in the equation connecting (3) and (5) even though, it is clear that that the observed rate of change was to be explained by the intensity of selection. This was no oversight. In 1924, there were no direct measurement of selection intensities. Rates of change, however could be estimated. As Haldane struggled to connect his models to data, selection intensities became predictions though, in the 1920s, their measurement did not seem forthcoming.

Part I starts with a haploid model. Let A and B be two phenotypes in a population in some generation and let pA:1B be their relative proportions. If, in the next generation, their proportions are pA:(1 - k)B, then Haldane defined "k" as the coefficient of selection. If $k = 1$, no B's survive, if $k = -\infty$, no A's survive.[15] Let the n-th generation have A and B (which can now be interpreted as alleles) in the proportion $u_n A$:1B. Assuming that generations do not overlap, if k is the selection coefficient,

$$u_{n+1} = \frac{u_n}{1-k}.$$

If $|k| << 1$, and $u_0 = 1$ (that is the A's and B's are initially present in the same proportion), the corresponding differential equation yields as a solution:

$$kn = ln\ u_n.$$

All other models, that is, those that allowed sexual reproduction, resulted in slower rates of change. In the case of non-sexual reproduction, Haldane observed, "speed must compensate to some extent for the failure to combine advantageous factors" (p. 22). This eventually became a common argument for the evolution of sex.

In a diploid model, let the two alleles, A and a be in the proportions $u_n A$:1a in generation n (that is, the proportion of the A and a gametes produced in generation n - 1 is u_n:1). Assuming random mating, non-overlapping generations

[15] Haldane's strategy for solving the models was simple. First, by calculating the change in the proportion of either type from one generation to the next, he wrote down a difference equation for the frequency of the alleles. These difference equations were characteristically too complicated for analytic solution and, in an age when computers were still to be invented, usually too cumbersome for practical use. However, Haldane found that he could usually convert the difference equation into a differential equation for two limiting cases, when the selection coefficient was very high or almost 0. The differential equations, though approximate, could be solved analytically and a useful result, that could be compared to experimental data, was obtained.

and complete dominance, the three zygotes survive to breed in the proportions $u_n^2 AA : 2u_n Aa$: (1 - k)aa. If $k \quad > 0$, the dominants have a selective advantage, if $k < 0$, the recessives are similarly favored. The change in the population is now described by the non-linear difference equation:

$$u_{n+1} = \frac{u_n(u_n + 1)}{u_n + 1 - k}.$$

If $|k| << 1$, and $u_0 = 1$, Haldane obtained the approximate relation:

$$kn = un + ln\ u_n - 1.$$

One of the numerical consequences of this equation eventually became a staple of textbooks of population genetics. Barrett (1895–1902) had described the case of the peppered moth (then known as *Ampidasys betularia*) the dominant melanic form of which first appeared in the Manchester area in 1848. By 1901 it had replaced the recessives almost completely. From Haldane's calculations, the selective difference was at least 1.5. As far as Haldane was concerned, this was not a "very intense degree of natural selection" (p. 26); for Fisher and others, however, it was unrealistically high.[16] There the matter lay for a generation, until Kettlewell [1956] finally measured the selection coefficient due to predation by birds and found it to be even higher. Even Haldane [1956] was not convinced that the coefficient could have always been quite that high. Nevertheless, this example, almost uniformly interpreted as a resounding success of evolutionary theory, percolated into the folklore of biology.[17] What it really shows is how difficult it has been to connect the models of population genetics to data collected from natural populations.

Besides these two cases, Haldane analyzed eleven other sets of models in the 1924 paper. He treated both autosomal and sex-linked loci and even considered the differences between models in which selection acted on only one or on both the sexes. What, in retrospect, appears to be the most innovative of these models, are those of various types of familial selection, for instance, between members of the same litter. Evolution slowed down considerably.[18] Most other models gave similar results. For instance, models of "certation" (gametic selection) allowed changes to be potentially as fast as in diploids. Even in those models in which selection potentially allowed faster changes, the rate was not much faster.

[16] See Provine [1971, 170].

[17] For a detailed account of the study of industrial melanism, and its incorporation into evolutionary studies, see Kettlewell [1973]. However, recent work has challenged the validity of Kettlewell's conclusions — see Majerus [1998].

[18] The standard approximate treatment (with $|k| << 1$ and $u_0 = 1$) now gave, instead of $kn = u_n + ln\ u_n - 1$, the equation:

$$\tfrac{1}{2}kn = u_n + ln\ u_n - 1.$$

Thus, the "species changes its composition at half the rate at which it would change if selection worked on the species as a whole, and not within families only" (p. 28). If the members of a family shared only one parent, the slowdown factor was 3/4 instead of 1/2, but the potential rate of change remained slower.

The slowness of the response to selection bothered Haldane. He worried whether the theory would satisfy the critical requirement: that selection must be able to "cause [a species] to change at a rate which will account for present and past transmutations" (p. 19). Change was most rapid if amphimixis (gametic fusion during sexual reproduction) could be avoided but this was not an option for most populations. Moreover, he observed, selection was particularly ineffective on most recessive traits. He suggested four ways of escaping from this conclusion: (i) self-fertilization or intense inbreeding; (ii) assortative mating; (iii) incomplete dominance; and (iv) heterozygote advantage. He also raised the possibility that isolation preferentially benefits the spread of recessives over dominants.

The first two of these factors — (i) self-fertilization and inbreeding and (ii) assortative mating — were taken up in Part II of this series of papers, incomplete dominance in Part III, and Haldane provided a general discussion of isolation in Part VI. If recessives were rare, self-fertilization and inbreeding greatly helped their spread. However, assortative mating had very little effect [Haldane, 1924b]. Incomplete dominance of an autosomal trait could help the spread of recessives; with sex-linked traits, it was of little help [Haldane, 1926]. Part III also included pioneering discussions of models of selection with multiple alleles at one locus; selection with tetraploidy; as well as a two locus model with linkage but no selection, where Haldane [1926] apparently independently rederived (with less elegance) some of the results of Robbins [1918b]. In Part IV, Haldane [1927a] turned to a demographic model with overlapping generations.[19] The mathematical tools required for the analysis of this model proved to be far more sophisticated than anything that had been necessary in the earlier papers: difference equations had to be replaced by integral equations. Most of Haldane's results had been previously obtained by Norton in 1910, and proved with much more rigor though they

[19] In passing, he provided a new — and simple — proof of Lotka's [1922] theorem of the stability of the normal age distribution.

had remained unpublished.[20] However, unlike Norton, Haldane remained focused on rates of change. The conclusion was disappointing: the temporal dynamics of change remained similar to the situation where generations did not overlap.

At this point Haldane seems to have given up all hope of escaping from the impasse that autosomal recessive factors cannot easily spread through a population without strong selection on recessives. Of the different schemes he had tried, only self-fertilization and inbreeding truly helped the recessives. Observing that most mutations were recessive to the normal type, in Part V, he turned to mutation as the source of recessive alleles. Fisher [1922] had analyzed the interaction of mutation with selection earlier and Haldane [1927b] also reached the conclusion that if the mutations are very rare, their survival depends almost entirely on stochastic factors. Haldane managed to calculate the probability of survival of a mutation. Assuming random mating in a diploid population, if a mutation is dominant, and it confers a slight selective advantage, k, the probability of its

[20]Norton's work was eventually published — see Norton [1928]. Norton died at the age of fifty. Years later, when R. F. Harrod [1951], who had been with Haldane at New College, Oxford, published a biography of John Maynard Keynes, he merely mentioned] Norton's work as an "application of probability theorems to certain problems in genetics" that had never been published (p. 188). Norton's sister, Betty, complained to Haldane (B. Norton to Haldane, January 31, 1951, Box 19, UCL) who wrote to Harrod, pointing out that Norton's work had eventually been published (Haldane to R. Harrod, February 2, 1951, Box 19, UCL) and:

> that the paper occupies some 45 pages, and that it gives the only rigorous proof of a great many notions which we are apt to take for granted in genetical work, but which in fact are only true if certain conditions are fulfilled. Besides this, it gives the methods for dealing with fairly complicated situations which sometimes arise. I suspect that its importance will grow as the application of mathematics to genetics advances. ...
>
> May I hope that in future editions you will do justice to a man who, though through ill health he did not achieve as much as was hoped, has certainly made a fundamental contribution to applied mathematics.

Harrod willingly, and apologetically acquiesced, submitting an emendation for future editions for Haldane's approval (R. Harrod to Haldane, February 3, 1951, Box 19, UCL).

Haldane also replied to Betty Norton (Haldane to B. Norton, February 2, 1951, Box 19, UCL). Norton's work, he claimed,

> is ... beyond doubt ... fundamental ... and I refer to it both in lectures and in a book which I am writing on the subject. In my own papers on the subject I had to refer to his unpublished work, since I had independently obtained some of his results, and of course knew his work. ...
>
> I have always felt that he may have thought that I 'jumped his claim', on the other hand the publication of his work was, if I remember, delayed for about seventeen years and as we only overlapped to a relatively small extent, I felt justified in going ahead with reference to 'the important unpublished work of H. T. Norton'. ...
>
> I very much hope that your brother did not feel that I had wronged him in any way, but you will realise that the position was difficult. I certainly feel that full justice should be done to his memory.

B. Norton wrote back reassuring Haldane on the last point (B. Norton to Haldane, February 3, 1951, Box 19, UCL).

survival is approximately equal to $2k$, and thus remains finite even in a large population. If the mutation is recessive, and the selective advantage that it confers is still k, the probability of its survival is $\sqrt{\frac{K}{N}}$, where N is the size of the population. Thus, the chance of survival decreases with the size of the population. Once again, inbreeding or self-fertilization brings this probability back to the level of the dominants.

If mutation occurred with an appreciable frequency, the situation was entirely different. A series of simple calculations showed that "any advantageous or not too disadvantageous factor will certainly be established" [1927b, 840] in the population. Haldane was impressed with the power of mutation:

> if selection acts against mutation, it is ineffective provided that the rate of mutation is greater than the coefficient of selection. Moreover, mutation is quite effective where selection is not, namely in causing an increase of recessives where these are rare. It is also more effective than selection in weeding out rare recessives provided that it is not balanced by back mutation of dominants. Mutation therefore determines the course of evolution as regards factors of negligible advantage or disadvantage to the species. [1927b, 842]

The only caveat was that it "can only lead to results of importance when its frequency becomes large" (p. 842). Haldane would never renege on this rather optimistic assessment of the evolutionary significance of mutations.[21] Six years later, when he returned to a systematic — but highly speculative — treatment of the topic, he invoked mutation to explain the disappearance of useless organs, recapitulation and the observation that the heterogametic sex is usually the male [Haldane, 1933].[22] However, Haldane's most important application of the results of this paper was his later use of the balance between selection and mutation to provide the first estimate of a human mutation rate [Haldane, 1935].

Three years later, at the end of 1929, he turned to a systematic analysis of isolated populations with immigration. Ten different sets of models were treated in Part VI [Haldane, 1930]. These models assumed that, in the isolated population, a new type is favored (with selection coefficient k), while migrants of the original type continue to enter the population. In each generation, the number of immigrants was l times the total isolated population. In each model unless k/l exceeded a certain critical value, as should perhaps be expected, immigration swamped out the original type. However, even if k/l did exceed that value, for the selected type to survive, it was sometimes necessary that its original frequency was high. When both these conditions were met, the population would reach an equilibrium composition with both types present. However, some of these equilibria were unstable: a fluctuation would drive one of the types to fixation.

[21] This is seen, for instance, in *The Causes of Evolution* [Haldane, 1932].

[22] Meanwhile, a consideration of the secondary effects of mutations led him into the burgeoning controversy, mainly between Fisher and Wright, over the evolution of dominance — Falk [2001].

Haldane had begun to appreciate how stochastic factors could influence the course of evolution. In Part VIII, perhaps the most interesting of the entire series of papers, he analyzed a two locus-two allele model with complete dominance and extended this discussion to a rudimentary analysis of n loci [Haldane, 1931b]. If there are two alleles at each locus, the genotypic space is an n-dimensional hypercube, with the genotypes at the vertices and edges connecting those genotypes that differ exactly at one locus. If fitnesses are assigned at random to these genotypes, there are at most 2^{n-1} stable points, that is, genotypes which were fitter than their nearest mutational neighbors. Haldane called populations in which such stable genotypes were fixed "metastable". The paper concluded with an intriguing remark: "It is suggested that in many cases related species represent stable types such as I have described, and that the process of species formation may be a rupture of the metastable equilibrium. Clearly such a rupture will be specially likely where small communities are isolated" [1931b, 141–142]. He had fully realized the power of stochastic factors, and this remark was an anticipation of Wright's "shifting balance theory" of evolution (see below). But it was no more than that — Haldane did not pursue the idea any further. In fact, Wright [1931] had just published the first version of that theory but that paper was unknown to Haldane when Part VIII was written.

Meanwhile, in Part VII, another intriguing paper, Haldane [1931a] analyzed the relation between competition and selection. Competition was modeled by what has come to be known as "truncation selection", acting on a normally distributed character: all individuals with a value higher than a specified value were eliminated from the population. Haldane showed that the intensity of selection may, under certain circumstances, decrease with an increase with the intensity of competition. This paper drove home the power of quantitative analysis on which Haldane had insisted right from Part I. As he put it, quantitative analysis showed that the assumption "often made that when competition is extremely intense at any stage in a life cycle, natural selection is bound to be intense also" [1931a, 131], was false. In fact, and even more surprising, "the intensity of selection may diminish and become negative at high rates of elimination" [1931a, 131]. Quantitative analysis thus exposed the limitations of qualitative argument.

In Part IX, Haldane [1932a] took up a one-locus model of rapid selection acting on a diploid model with full dominance. The model was the same as the one he had analyzed in Part I except, now, he set out to find an approximate solution of equation (4.1) without assuming that $|k| << 1$.[23] The analysis permitted a quantitative evaluation of the ecologist, Charles Elton's [1927] intriguing suggestion that episodic intense selection, such as that due to plagues or famines, was more effective in changing the composition of a population than less intense selection acting every generation. Haldane found that this was true if selection favored the dominants, but not if it favored the recessives. Once again, quantitative analysis revealed intricacies that no amount of qualitative argument could have shown.

[23]The mathematical difficulty of finding such a solution was formidable, and Haldane [1932c] published the formal analysis separately.

The final part of the "Mathematical Theory" consisted of a collection of theorems about artificial selection, more important for breeders than for an understanding of evolution in nature [Haldane, 1934]. With its publication ended the most sustained research program of Haldane's career. It had begun in 1922 and had dominated his work for over a decade. The "Mathematical Theory" was Haldane's most important contribution population genetics.[24]

To a first approximation, at least, it is clear what the "Mathematical Theory" had achieved. In 1922, when Haldane had embarked on his project, as Wright [1968, 3], has aptly observed, "the precise effect of selection on the composition of a Mendelian population had been presented in only the simplest cases". The "Mathematical Theory" was the first comprehensive examination of that effect. In 1922, it could still be doubted that natural selection *alone*, acting on blind variation produced by mutations, could account for all of evolutionary change and, in particular, for the observed rates of evolutionary change. Alternatives such as orthogenesis were still viable candidates as mechanisms of evolutionary change. The "Mathematical Theory", along with the work of Fisher and Wright, and the work emerging from Chetverikov's group in the Soviet Union, removed these alternatives and put the discussion of the power of a natural selection on an entirely new level of rigor: from now on, verbal intuitions had to be supplemented by quantitative reasoning. There would later be worries about the power of natural selection. Wright would initiate these worries in 1932, as he began to emphasize the role of drift in evolution. Haldane, himself, had realized that natural selection, unless very strong, could also prove to be ineffective in some circumstances, for instance, in the spread of recessive alleles in a random mating population. Later, there was an even stronger challenge to the basic thrust of the "Mathematical Theory", when Mayr would suggest that quantitative reasoning is of little use in evolutionary theory. But these later disputes would occur in a permanently transformed conceptual space. This was what the "Mathematical Theory", and the ongoing work of Fisher and Wright in the 1920's, had achieved.

Nevertheless, unlike the work of Fisher and Wright, there is no central theme to the "Mathematical Theory", no grand hypothesis of *the* preferred mode of evolutionary change. Fisher insisted on the importance of natural selection acting mainly upon single genes in large panmictic populations. Wright would eventually emphasize drift, isolation and population structure though he did not deny the importance of selection. Fisher put his faith in what he called the "fundamental theorem of natural selection", that the increase in fitness of a population through selection would be equal to its additive genetic (genic) variance in fitness. Wright

[24]Fifteen years after the completion of the series, in 1949, when the mathematician, J. Neyman began composing a short treatise on probability and statistics, with a chapter devoted to genetics, he asked Haldane for his most important papers (J. Neyman to Haldane, November 14, 1949, Box 19, UCL). Haldane's response was unequivocal: "I think my most important work is in the Transactions of the Cambridge Philosophical Society from 1924 to 1936" (Haldane to J. Neyman, November 22, 1949, Box 19, UCL). The details of the statement are inaccurate — only Part I of the "Mathematical Theory" had appeared in the *Transactions*; the next eight appeared in the *Proceedings*.

would finally codify his ideas into the three-phase "shifting balance theory" (see below).[25] In sharp contrast, the "Mathematical Theory" appears almost mundane. Models of selection are introduced, their virtues and vices discussed, and then almost cast aside as other models are explored. The few general claims, most notably, the skepticism about the power of selection to spread recessive genes and the possible importance of mutations, are carefully presented with many caveats and qualms. Haldane carefully discusses the known experimental data, and was more concerned with connecting models to data than either Wright or Fisher were during this period. As he pointed out, himself, Part I included a discussion of "all the then known types of single-factor Mendelian inheritance" [1927b, 843]. However, the "Mathematical Theory" does not even contain a systematic discussion of the relevance of the models to all the experimental results that had accumulated during this period. The latter lacuna, at least, was removed in the *Causes of Evolution*.

4 *THE CAUSES OF EVOLUTION*

In January 1931 Haldane delivered a series of lectures at the Prifysgol Cymru at Aberystwyth, Wales, entitled "A Re-examination of Darwinism". The text of these, along with an Appendix, became Haldane's best-known book, *The Causes of Evolution*.[26] Haldane started with the unquestionable fact of evolution, "the descent from living beings in the past of other widely different living beings" [1932, 4]. All that remained debatable were its possible causes.

The first chapter considered five causes that had been historically suggested "for the deeper transformations of the geological record" [1932, 11]: (a) basically random inheritable variation; (b) environmentally induced inheritable variation; (c) variation due to "internal causes"; (d) variation due to hybridization; and (e) selection. Though Darwin had coupled selection with both random and acquired inheritable variation, experimental evidence had largely ruled out the latter. Those who believed in natural selection as a major cause of evolution were thus left with selection acting on random (or blind) inheritable variation. Haldane's series of papers had provided a quantitative basis for neo-Darwinism. It remained to provide a full biological interpretation of the new theory, to show how the new genetics permitted the construction of a complete theory of evolution from only causes (a) and (e). This was the major contribution of *Causes*.

By the late 1920's, the so-called Lamarckists who advocated the inheritance of acquired characters were rare. In *Causes*, Lamarckism is dismissed because of its experimental failures (pp. 130–138). The other alternatives to neo-Darwinism

[25] See Crow [1990] for a particularly lucid introduction to the shifting balance theory. It is reproduced in Sarkar [1992].

[26] In spite of its reputation for being well-written, *Causes* is hard to follow. Arguments are often imprecise and the narrative disjointed. It reads like a transcript of lectures that could have profited from more careful editing. One of the most humorous of the editorial lapses occurs in the list of references where all of Haldane's works are attributed to Charlotte.

such as saltationism and orthogenesis (that is, cause [c]) were more important. In the 1920s, saltationism came with an illustrious pedigree, at least among geneticists.[27] Though Haldane emphasized neo-Darwinism, he remained ambivalent about saltationism and regarded it as a live possibility, at least for plant speciation.

In the case of orthogenesis, invoking selection to explain the usual cases of evolutionary progress through adaptation was straightforward. However, Haldane also deployed a different argument against orthogenesis. He denied the ubiquity of evolutionary progress and, thus, the possibility of any inner perfecting principle:

> Degeneration is a far commoner phenomenon than progress. It is less striking because a progressive type, such as the first bird, has left many different species as progeny, while degeneration often leads to extinction, and rarely to a widespread production of new forms. Just the same is true with plants. ... Certainly the study of evolution does not point to any general tendency of a species to progress. The animal and plant community as a whole does show such a tendency, but this is because every now and then an evolutionary advance is rewarded by a very large increase in numbers, rather than because such advances are common. But if we consider any given evolutionary level we generally find one or two lines leading up to it, and dozens leading down. (p. 153)

According to Haldane, all talk of progress "represents rather a tendency of man to pat himself on the back[;] ... [t]he change from monkey to man might well seem a change for the worse to a monkey" (p. 153).

Chapter II summarized what was known about intra-specific variation at the levels of both genes and chromosomes, that is, the work of the Morgan school, which was given an evolutionary interpretation. Chapter III turned to variation between species. What intrigued Haldane was that species could differ by entire sets of chromosomes. This automatically suggested speciation through hybridization and allopolyploidy, and Haldane provided the example of *Primula floribunda* and *Primula verticillata* being hybridized to form *Primula kewensis* (pp. 67–69). Allopolyploidy was common among plants, though rare among animals if it occurred at all. Ultimately, Haldane concludes:

> interspecific differences are of the same nature as intervarietal. But the latter are generally due to a few genes with relatively large effects, and rarely to differences involving whole chromosomes or large parts of them. The reverse is true of differences between species. The number of genes involved is often great, and cytologically observable differences common. It is largely these latter which are the causes of interspecific sterility. (p. 82)

[27] Haldane was particularly impressed by Willis's [1922] *Age and Area* but also mentioned de Vries and Bateson. Though he did not mention them, Goldschmidt and, to a lesser extent, Morgan also adhered to forms of saltationism. See Mayr [1982, 540–550].

Having thus integrated the experimental results of genetics into the framework of evolution, Chapter IV begins the analysis of selection. It provides a systematic biological interpretation of the "Mathematical Theory", pooling together the more important examples that had been scattered through its ten parts. But, beyond that, it reveals a more sophisticated view of natural selection than the earlier work. In particular, in both this chapter and the next, Haldane's biochemical expertise (recall that, professionally, he was a biochemist[28]) is brought to bear on genetics to an extent that neither he, nor anyone else, had ever attempted before. For instance, he notes the possibility of epistasis: "It is important to realise that the combination of several genes may give a result quite unlike the mere summation of their effects one at a time" (p. 96). He immediately a suggests a biochemical explanation: "[epistasis] is obviously to be expected if genes act chemically" (p. 96).[29]

Haldane also notes the possibilities of pleiotropy and neutral or nearly-neutral genes. With respect to pleiotropy, "[o]ccasionally a single gene might produce simultaneous and harmonious changes in many [characters] at one, but this is not generally the case with new mutants, though some such genes, being almost harmless, are not eliminated, and account for much of the variation in natural populations" (p. 103). Haldane was keenly aware of how this affected the rate of evolution: "Evolution must have involved the simultaneous change in many genes, which doubtless accounts for its slowness" (p. 103). However, neutral or nearly neutral mutations may have enabled rapid evolutionary change. "If the only available genes produce large changes, disadvantageous one at a time", Haldane explained,

> then it seems to me probable that evolution will not occur in a random mating population. In a self-fertilized or highly inbred species it may do so if several mutations useful in conjunction, but separately harmful, occur simultaneously. ... But when natural selection slackens, new forms may arise which would not survive under more rigid competition, and many ultimately hardy combinations will thus have a chance of arising. ... This seems to have happened on several occasions when a successful evolutionary step rendered a new type of organism possible, and the pressure of natural selection was temporarily slackened. (pp. 104–105)

Hybridization was another possible source of rapid evolutionary change. The chapter ends with a discussion of the relative importance of mutation and selection. Haldane's ultimate conclusion is a little more negative than in the "Mathematical

[28] Sarkar [1992a] discusses Haldane's biochemical work.

[29] *Primula sinensis* provides an example: in it, "a dark stem (recessive) is associated with no great change in colour of acid-sapped (red and purple) flowers. But blue (recessive flowers, which have a neutral sap, when growing on a dark stem are mottled. The same recessive dark stem genes, along with genes for a green stem, give plants which will not set seed, though they give good pollen" (p. 96).

Theory": "we cannot regard mutation as a cause likely by itself to cause large changes in a species" (p. 110).

The most innovative discussions in Chapter V were of of altruism and of the conflict between competition and selection (which went beyond Part VII of the "Mathematical Theory"). Both were motivated in part by political concerns. Haldane was concerned to expose the "poisonous nonsense which has been written on ethics in Darwin's name" (p. 119). These writings were based on a

> fallacy . . . that natural selection will always make an organism fitter in its struggle with the environment. This is clearly true when we consider the members of a rare and scattered species. It is only engaged in competing with other species, and in defending itself against inorganic nature. But as soon as a species becomes fairly dense matters are entirely different. Its members inevitably begin to compete with one another. . . . And the results could be biologically advantageous for the individual, but ultimately disastrous for the species. The geological record is full of cases where the development of enormous horns and spines (sometimes in the male sex only) has been the prelude to extinction. It seems probable that in some of these cases the species literally sank under the weight of its own armaments. (pp. 119 -120)

There is a clear recognition of the potential conflict between various levels of selection in this passage.[30]

A less explicit but more influential recognition of different levels of selection is to be found in the other major innovation of this chapter, the account of altruism:

> It can be shown mathematically that in general qualities which are valuable to society but usually shorten the live of their individual possessors tend to be extinguished by natural selection in large societies unless these possess the type of reproductive specialisation found in social insects. This goes a long way to account for the much completer subordination of the individual to society which characterises insect as compared to mammalian communities. (p. 130)

The mathematical argument from the Appendix (pp. 207–210) will be reconstructed below. Haldane's conclusion was that: while he "doubt[s] if man contains many genes making for altruism of a general kind, ... we do probably possess an innate predisposition for family life. ... For in so far as it makes for the survival of one's descendants and near relations, altruistic behaviour is a kind of Darwinian fitness, and be expected to spread through natural selection" (p. 131).

Despite the spirited defense of natural selection, at the end of the book, Haldane remained cautious about its power. Two other factors must be included in

[30] Mayr [1992] has read this passage to indicate that Haldane, like many others in the 1920's, was assuming that selection is for the benefit of the species rather than the individual. However, what Haldane seems to be saying is that individual selection relative to other individuals in a population need not give any indication of the rate of change of the population size.

a general account of evolution: (i) discontinuous changes (for instance through hybridization); and (ii) the mutations to generate the necessary variation. As Haldane put it:

> if we come to the conclusion that natural selection is probably the main cause of change in a population, we certainly need not go back completely to Darwin's point of view [though without cause (b)]. In the first place we do have reason to believe that new species may arise quite suddenly, sometimes through hybridisation, sometimes perhaps by other means. Such species do not arise, as Darwin thought, by natural selection. When they have arisen they must justify their existence before the tribunal of natural selection, but that is a very different matter. ...
>
> Secondly, natural selection can only act on the variations available, and these are not, as Darwin thought, in every direction. In the first place, most mutations lead to a loss of complexity. ... This is probably the reason for the at first sight paradoxical fact that ... most evolutionary change has been degenerative. But further, ... mutations only seem to occur along certain lines, which are very similar in closely related species, but differ in more distant species. (pp. 138–140)

For the next generation of population geneticists, the Appendix of *The Causes of Evolution* was at least as important as the text. In the Appendix, Haldane summarized the mathematical results obtained by Fisher, Wright and himself. Given that Fisher had uncharitably altogether ignored both Haldane's and Wright's work in the *Genetical Theory*, and Wright did not write a general review or book during this period, this Appendix becane the only compendium of mathematical genetics.

Besides the summary, Haldane presented a new analysis of altruism. Explaining altruism within families, or in insect societies where many individuals were genetic clones of each other, was relatively straightforward. But Haldane went further: consider a group of N diploid individuals, mating at random with respect to the altruism trait in a one locus-twoalleles (A, a) model with complete recessivity. Then the relative proportions of the three possible genotypes are $u_n^2 AA : 2u_n Aa$: $1aa$. Let the recessives, aa, leave (1 -k) times as many progeny as the AA or Aa ($0 < k < 1$). Thus the altruism trait decreases individual fitness. However, let the presence of a proportion x of altruistic (aa) individuals result in an increase in the probably progeny of all individuals by a factor $(1 + Kx)$. If it is assumed that a group consisting entirely of altruistic individuals should grow, $K > k$. In the next generation, since $x = \frac{1}{(1+u_n)^2}$, the size of the group will have increased to $N\left[1 + \frac{K}{(1+u_n)^2}\right]$. The recessive allele number will thus have changed from $\frac{N}{1+u_n}$ to $N\left[1 + \frac{K}{(1+u_n)^2}\right]\frac{u_n+1-k}{(1+u_n)^2}$. Neglecting kK, the change in the number of a alleles, Δx is then given by:

$$\Delta x = \frac{N(\frac{K}{(1+u_n)} - k)}{(1+u_n)^2}.$$

This means that the number of recessive alleles (a) will increase if $1 + u_n < \frac{K}{k}$, that is, $\Delta x > 0$.[31] Therefore, recessive alleles for altruistic behavior can increase in proportion only if they are already common. However, since u_n will also increase, this process will end. Similar results are obtained if the altruism trait is dominant. Haldane concluded that "the biological advantages of altruistic conduct only outweigh the disadvantages if a substantial proportion of the tribe behave altruistically. If only a small fraction behaves in this manner, it has a very small effect on the viability of the tribe, not sufficient to counterbalance the bad effect on the individuals concerned" (p. 209).

Moreover, for altruism to spread, the proportion of altruists need not be great if $\frac{K}{k}$ is large. In fact, for a recessive altruism trait, if $\frac{K}{k} > \sqrt{N}$, a single altruistic individual is sufficient. Thus, for small groups there always is effective selection for altruism. However, in large groups, according to Haldane, "the initial stages of the evolution of altruism depend not on selection, but on random survival" (p. 209). This had some bearing on possible human altruism:

> If any genes are common in mankind which promote conduct biologically disadvantageous to the individual in all types of society, but yet advantageous to society, they must have spread when man was divided into small endogamous groups. As many eugenists have pointed out, selection in large societies operates in the reverse direction. ... I find it difficult to suppose that many genes for absolute altruism are common in man (pp. 209–210).

5 THE AFTERMATH

There is one relatively uncontroversial sense in which the joint work of Fisher, Haldane, and Wright was both similar and important: it provides a quantitative account of the effects of selection acting on Mendelian populations. With the publication of *Causes* in 1932, one chapter in the history of evolutionary theory — and the only chapter to which Haldane made decisive contributions — came to a close. Whereas Fisher and Wright continued to refine their respective theories of evolution until their deaths, Haldane largely turned aside from evolutionary theory in the 1930's and his subsequent contributions to that theory were sporadic.

What is lost in this broad picture is the extent to which *Causes* differed from Fisher's and Wright's work. Fisher [1930] argued that evolution proceeded by continuous selection acting upon single genes, each being only subjected to a small selective pressure, in large populations. He showed that the frequencies of each gene would reach an equilibrium but, and this ultimately was the source of progressive change in Fisher's theory, the environment deteriorated. So selection had to continue, as the species struggled to keep up with the environment.

[31] Haldane [1932, 209] reports the relation as $1 + u_n > \frac{K}{k}$ which is incorrect. However, his discussion (pp. 209–210) is based on the correct result!

Though Wright did not write a book during this period, at the end of his long paper from 1931, he presented his theory of evolution in detail:

> Evolution as a process of cumulative change depends on a proper balance of the conditions, which, at every level of organization — gene, chromosome, cell, individual, local race — make up for genetic homogeneity or genetic heterogeneity of the species. ... The type and rate of evolution ... depend upon the balance among the evolutionary pressures considered here. In too small a population ... there is nearly complete fixation [of an allele at each locus], little variation, little effect of selection and thus a static condition modified occasionally by chance fixation of new mutations leading inevitably to degeneration and extinction. In too large a freely interbreeding population ... there is great variability but such a close approach to complete equilibrium of all gene frequencies that there is no evolution under static conditions. ... In a population of intermediate size ... there is continual random shifting of gene frequencies and a consequent shifting of selection coefficients which leads to a relatively rapid, continuing, irreversible, and largely fortuitous, but not degenerative series of changes, even under static conditions. The rate is rapid only in comparison with the preceding cases ... being limited by mutation pressure. Finally in a large population, divided and subdivided into partially isolated local races of small size, there is a continually shifting differentiation among the latter ... which inevitably brings about an indefinitely continuing, irreversible, adaptive, and much more rapid evolution of the species. Complete isolation in this case, and more slowly in the preceding, originates new species differing for the most part in nonadaptive respects but is capable of initiating an adaptive radiation. [Wright, 1931, 158–159]

Wright clearly preferred the last possibility, and there, in outline, were the rudiments of what Wright eventually refined as his three-phase "shifting balance theory of evolution".

In contrast to Fisher and Wright, Haldane consistently refused to endorse any single factor as *the cause* of evolution. In this sense, Haldane did not have a *unified theory* of evolution. This attitude stems from Haldane's catholic interest in all details of evolutionary change, as revealed through genetic and biochemical work. Evolution was too diverse for Haldane to endorse a particular theory. Fisher conceived of the theory of natural selection as an analog of the most universal of physical theories, the kinetic theory of matter. Wright came to believe that all important evolutionary change was dominated by population structure, size, patterns of isolation and, perhaps only to a lesser extent, selection. Haldane refused to endorse either of these alternatives as being *the* cause of evolution. Though he did not present any alternative he did something which, in retrospect, was perhaps even more important. He systematically explored every evolutionary

context that could be effectively modeled. What was surprising — and reassuring to Haldane — was that natural selection was important in so many of them. But natural selection was subtle: it worked in a bewildering variety of ways. Evolution was complex: for Haldane, there was no single cause of evolution.

Five years later, in the altered conceptual landscape Dobzhansky [1937] was the first to apply the new genetics to the problem of speciation and his treatment was a significant advance over Haldane — particularly in the incorporation of the conceptual apparatus being developed by the naturalists. He coined the term "isolating mechanism" to denote the factors that maintain reproductive barriers between populations. He included both genetic factors and geographical isolation among possible isolating mechanisms. Huxley [1940] collected together major contributions to systematics that led to the breakdown of the concept of the species as a type. Mayr [1942] soon followed with his *Systematics and the Origin of Species* in which he collected an extraordinary body of data showing how geographical isolation was critical for speciation. Mayr called this mechanism "allopatric speciation" which he distinguished from "sympatric speciation", which would not require geographic isolation. The latter was supposed to be rare enough to be unimportant in the course of evolution. Mayr ignored the fact that Wright had been emphasizing the importance of isolation since 1931. What Mayr, in effect, had actually achieved was to provide data from the field in support of Wright's emphasis. This is how Dobzhansky read Mayr in 1955 (see below). Mayr, however, came to think rather differently about what he had achieved.

6 HISTORICAL RECONSTRUCTIONS

To return to the two reconstructions of the history of evolutionary theory in the early twentieth century, in the early 1970's, Provine's original views reflected what, with very few exceptions, had been the scientific consensus for a generation. Huxley put it this way in 1942:

> Darwinism to-day ... still contains an element of deduction, and is none the worse for that as a scientific theory. But the facts available in relation to it are both more precise and more numerous [than in Darwin's time], with the result that we are able to check our deductions and to make quantitative prophecies with much greater fullness than was possible to Darwin. This has been especially notable as regards the mathematical treatment of the problem, which we owe to R. A. Fisher, J. B. S. Haldane, Sewall Wright, and others. [1942, 21]

In 1954, the naturalist, P. M. Sheppard, was more blunt about the role of Fisher, Haldane, and Wright: "The great advances in understanding the process of evolution, made during the last thirty years, have been a direct result of the mathematical approach to the problem adopted by R. A. Fisher, J. B. S. Haldane, S. Wright, and others" [1954, 201]. However, the "mathematical theories give no

information on the conditions found in nature, but only show in what circumstances different factors can be of importance in evolution" (p. 201). The role of field work was to discover those conditions.

One year later, at a Cold Spring Harbor symposium on quantitative biology, Dobzhansky stated it equally forcefully. For him, population genetics was central to the study of evolution [1955, 12]. Moreover:

> The foundations of population genetics were laid chiefly by mathematical deduction from basic premises contained in the works of Mendel and Morgan and their followers. Haldane, Wright, and Fisher are the pioneers of population genetics whose only research equipment was paper and ink rather than microscopes, experimental fields, *Drosophila* bottles, or mouse cages. Theirs is theoretical biology at its best, and it has provided a guiding light for rigorous quantitative experimentation and observation. [1955, 13–14]

And, in 1966, George C. Williams repeated the consensus: "The study of adaptation [which he took to be the core of evolutionary theory] has already had its Newtonian synthesis [which] is the genetical theory of natural selection, a logical unification of Mendelism and Darwinism that was accomplished by Fisher, Haldane and Wright more than thirty years ago" [1966, 20].

Perhaps unsurprisingly, Provine started from the same position:

> The origins of population genetics is perhaps best understood as a product of the conflict between two views of evolution which were eventually synthesized. On the one side was Darwin's belief in gradual evolution produced by natural selection acting upon small continuous variations. On the other was Galton's belief in discontinuous evolution. ... The theoretical foundations [of the synthetic theory], sometimes termed 'classical' population genetics, were laid between 1918 and 1932 by R. A. Fisher, J. B. S. Haldane, and Sewall Wright. [1971, ix–x]

However, Provine's genetical account underwent a significant change in the late 1970's, as he embarked on a biography of Wright with Wright's active collaboration.[32] Provine encountered the unresolved dispute between Fisher and Wright about what mechanism had *historically* been the most significant agent of evolutionary change on Earth. As Provine [1986] records, the dispute had become vitriolic, partly because the antagonism between Wright and Fisher became personal in the 1930s. There was little prospect of resolving the dispute using the experimental data or analytic techniques of the late 1970's. Provine chose to emphasize the significance of the dispute and, again not surprisingly, emerged as a partisan for Wright, ending his biography with a bold claim: "I predict that historians and biologists of the twenty-first century will look upon Wright as the single most influential evolutionary theorist of this century" (p. 499).

[32]See Provine [1986, xiii].

In the dispute between Fisher and Wright, Provine sided with Wright. But there was the broader dispute, that between the genetical and naturalist narratives mentioned at the beginning of this paper. In that dispute, both Fisher and Wright and, for that matter, Haldane, Kimura and the other theoretical population geneticists, are jointly pitted against Mayr and his followers. By emphasizing the significance of the Wright-Fisher dispute in the history of evolutionary theory, Provine implicitly denies Mayr's reconstruction. There will be much more of that below but, meanwhile, the relevant question here is why Provine's later accounts also ignores Haldane. Part of the answer may lie in Haldane's consistent refusal to take sides in the dispute between Fisher and Wright: each of them found Haldane to be closer in position to the other. Provine, following Wright, presumably placed Haldane with Fisher and hence doubted his significance.[33] But the main reason is Provine's fundamental decision to present the Fisher-Wright dispute as the most central issue in the history of evolutionary theory. A perhaps inevitable consequence of this decision, in spite of all the achievements recorded earlier in this paper, is the elision of Haldane's role in that history.

To turn to the naturalist account, already in the 1950s, the genetical account began to be challenged. The first to explicitly challenge it was C. H. Waddington. In 1952, in Oxford, at the Symposium of the Society for Experimental Biology, Waddington argued that the prestige of the mathematical work of Fisher, Haldane and Wright was unwarranted. It

> did not achieve either of the two results which one normally expects from a mathematical theory. It has not, in the first place, led to any noteworthy quantitative statements about evolution. The formulae involve parameters of selective advantage, effective population size, migration and mutation rates, etc., most of which are still too inaccurately known to enable quantitative predictions to be made or verified. But even when this is not possible, a mathematical treatment may reveal new types of relation and of process, and thus provide a more flexible theory, capable of explaining phenomena which were previously obscure. It is doubtful how far the mathematical theory has done this. Very few qualitatively new ideas have emerged from it [1953, 186].

Waddington [1957] republished these remarks verbatim in 1957, in *The Strategy of the Genes*. For him, the only contributions of mathematical population genetics were (i) the demonstration that ordinary Mendelian genes "would respond to the process of natural selection" (p. 61); and (ii) the demonstration that continuous variation can be accommodated within a Mendelian framework, that is, Fisher's work from 1918.

[33] Moreover, almost all the population genetic models that Haldane had ever constructed were deterministic. They assumed large (in principle, infinite) populations. They were models in which selection would generally prevail over population structure; certainly, they had no room for stochasticity. Here, even from a mildly Wrightian point of view, Haldane and Fisher clearly belonged together.

However, the critique of theoretical population genetics that became most influential in the 1960's, and continues to be influential today, was due to Mayr. Mayr had read Waddington's critical remarks after they were republished in *The Strategy of the Genes* in 1957, and had agreed with them.[34] At the 1955 Cold Spring Harbor symposium at which Dobzhansky had delivered his ringing endorsement of theoretical population genetics, Mayr had distinguished between the approaches of the mathematical population biologists and "field naturalists". But, in sharp contrast to Sheppard's 1954 assessment, for Mayr, the field naturalists were not simply providing data for the theorists. Rather, they were supposed to be have a different and more sophisticated, though not entirely satisfactory, view of the evolutionary process. Mayr elaborated on this point: what was wrong with the mathematical population geneticists was that their analysis was limited to "simple Mendelian characters or simple frequencies of genes in natural populations" whereas what was required was the consideration of "entire genetic complexes" (p. 327). The field naturalists were also at fault for not sufficiently integrating genetics into their work. However, according to Mayr, what was needed was physiological genetics, rather than mathematical population genetics. Mayr did not directly criticize Fisher, Haldane and Wright. Rather, he did not mention them and, instead, endorsed the work of Dobzhansky, King, Thoday, Wallace and others as being the most relevant to the creation of a "new population genetics" which incorporated the "integration of genotypes" (p. 332). The result, he grandiosely claimed, was "a theory of relativity in the field of population genetics" (p. 333).

Presumably, Mayr's self-acknowledged inability to follow the mathematical work of Fisher, Haldane and Wright's models had prevented him from being more critical in 1955. Waddington, however, was reputed to have the necessary mathematical skills. In 1959 Mayr, emboldened by Waddington's remarks, presented his criticisms of the mathematical population geneticists more directly. That year, the Cold Spring Harbor Symposium on Quantitative Biology was devoted to "Genetics and Twentieth Century Darwinism" to mark the centenary of the publication of Darwin's *Origin*. Mayr, in a provocative address, "Where Are We?", fully challenged the orthodox view of the role of theoretical population genetics in evolutionary theory.

Mayr distinguished three periods in the history of evolutionary genetics:

(i) a "Mendelian" period from 1900 to 1920 allegedly dominated by the view that evolution took place through mutations;

(ii) a period of "classical population genetics", from 1920 to the late 1930's, during which the

> emphasis . . . was on the frequency of the genes and on the control of this frequency by mutation, selection, and random events. Each gene was essentially treated as an independent unit favored or discriminated against by various causal factors. In order to permit

[34] Ernst Mayr (personal communication).

> mathematical treatment, numerous simplifying assumptions had to be made, such as that of an absolute selective value of a given gene. The great contribution of this period was that it restored the prestige of natural selection ... and that it prepared the ground for a treatment of quantitative characters. Yet, this period was one of gross oversimplification. Evolutionary change was essentially presented as an input or output of genes, as the adding of certain beans to a beanbag and the withdrawing of others. (p. 2)

"Beanbag genetics", according to him, had outlived its utility and had to be replaced by

(iii) the "newer population genetics", starting in the late 1930's, and "characterized by an increasing emphasis on the interaction of genes" (p. 2), leading to the "theory of relativity" that he had pompously announced in 1955.

The period of beanbag genetics, according to Mayr, had been dominated by the "mathematical analyses and models" of Fisher, Haldane and Wright. Mayr put his challenge to them bluntly: "These authors, although sometimes disagreeing with each other in detail or emphasis, have worked out an impressive mathematical theory of genetical variation and evolutionary change. But what, precisely, has been the contribution of this school to evolutionary theory, if I may be permitted to ask such a provocative question?" (p. 2). Explicitly endorsing Waddington's [1957] criticisms, Mayr attempted to provide his answer:

> It seems to me that the main importance of the mathematical theory was that it gave mathematical rigor to qualitative statements long previously made. It was important to realize and to demonstrate mathematically how slight a selective advantage could lead to the spread of a gene in a population. Perhaps the main service of the mathematical theory was that in a subtle way it changed the mode of thinking about genetic factors and genetic events in evolution without necessarily making any startlingly novel contributions. (p. 2)

Mayr concluded: "I should perhaps leave it to Fisher, Wright, and Haldane to point out themselves what they consider to be their major contributions" (p. 2).

Fisher, who died in 1962, never responded to Mayr's remarks. However, Wright's response was immediate. Wright was present at Cold Spring Harbor for the 1959 symposium though he did not present a paper. However, when asked to review the proceedings for the *American Journal of Human Genetics*, Wright only listed the other papers in the volume while devoting the entire review to an analysis of Mayr's contentions.[35] He challenged Mayr's interpretation of the history of genetics. Between 1900 and 1920, he pointed out, that many geneticists had been

[35] For more detail, see Provine [1986, 481–484].

avid Darwinians — the mutationists were by no means dominant. Moreover, during Mayr's "classical period", much of what population genetics had achieved was exactly what Mayr had attributed to the "newer population genetics". Contrary to Mayr's accusation of oversimplification, work in mathematical population genetics in the late 1920s and 1930s had incorporated allelic interactions (degrees of dominance), epistasis, frequency-dependent selection, and other factors. Most importantly, Wright correctly insisted that he had emphasized gene interactions throughout his life and that his parameter for selective value (or fitness) had "always been defined as applying to a *total* genotype in the system under consideration, thus involving whatever interaction effects there may be among the component genes" [Wright, 1960, 369].

Nevertheless, in 1963, Mayr republished his criticisms in *Animal Species and Evolution*. In a chapter entitled "The Unity of the Genotype", he claimed:

> The procedure of classical Mendelian genetics, of studying each gene locus separately and independently, was a simplification necessary to permit the determination of the laws of inheritance and to obtain basic information on the physiology of the gene. When dealing with several genes, the geneticist was inclined to think in terms of their relative frequencies in the population. The Mendelian was apt to compare the genetic composition of a population to a bag of colored beans. Mutation was the exchange of one kind of one kind of bean for another. This conceptualization has been referred to as 'beanbag genetics.' ... Work in population and developmental genetics has shown, however, that the thinking of beanbag genetics is in many ways quite misleading. To consider genes as independent units is meaningless from the physiological as well as the evolutionary viewpoint. Genes not only *act* ... but also *interact*. (p. 263)

Mayr ignored Wright's response altogether.

Meanwhile, Haldane had briefly commented on Waddington's criticism in 1953. Writing the Foreword for the proceedings of the Oxford symposium at which Waddington had first stated his criticisms, Haldane commented:

> I shall first take advantage of having the last word by defending myself, in my capacity as mathematical geneticist, against Waddington's criticism that 'very few qualitatively new ideas have emerged from it'. I think that some of them have emerged so completely that their origin is forgotten. Darwin apparently [thought] that variation could only [arise] by the effects of environmental differences in each generation. Fisher first showed clearly that in large populations it diminished with extreme slowness apart from the effects of natural selection, and Hagedoorn that it would diminish at a measurable rate in small populations. Wright then produced the qualitatively new idea of 'drift' or random evolution in small populations . ... Penrose and Haldane first showed

> that some genes which seriously lower human fitness are only kept in being by mutation. ... Another idea which seems to me qualitatively new is that, because a selective advantage of the heterozygote over both homozygotes will preserve both of a pair of [alleles] indefinitely, we may expect most of the genetic variation in an outbred population to be due to genes with this peculiar property. Such ideas as these pass rapidly from being mathematical theorems to being common sense. Other ideas with a superficial appeal to common sense, for example that dominants must oust recessives, appear to require mathematical disproof to prevent their spread. Mathematical biologists have also suggested what other biologists should measure. (pp. ix–x)

Haldane went on to discuss examples. With a display of striking intellectual irresponsibility, Waddington ignored these remarks when he republished his criticism verbatim in 1957.

But Waddington was peripheral to the development of evolutionary theory, and his criticism (by itself) would have probably had little influence without their endorsement by Mayr who, along with Dobzhansky, Simpson, and a few others, was at the forefront of evolutionary theory in the 1950s and 1960s. In 1964, Haldane took up Mayr's challenge and provided a defense of mathematical population genetics. The result is a methodological classic, an ultimate defense of reductionism in evolutionary biology. Haldane started with the passage from *Animal Species and Evolution* that was quoted above, and Mayr's "provocative question" from 1959. He admitted that the models of mathematical population genetics made simplifying assumptions. Then he proceeded, by explicitly listing achievements, to argue for the importance of even single locus models, basically the only ones to which the epithet "beanbag genetics" could be attributed with justice:[36]

(i) Wright and he had shown around 1930 that, if a species had a population of over a hundred thousand, the mutation rate does not determine the rate of evolution, even if selection is weak;

(ii) using the same basic model, Haldane had been able to provide the first estimate of a human mutation rate when there is an equilibrium between mutation and selection. How soon such an equilibrium is reached in different genetic contexts could only be known from a mathematical analysis;

(iii) Mayr [1963, 191] had mis-stated the situation even with respect to some developments whose importance he admitted. Mayr had claimed that Haldane [1932] had assumed only small selective differences in his work; that, to the contrary, in the case of industrial melanism in the peppered moth, *Biston betularia*, a selective difference of 30 -50% had been observed; and that Haldane had only reached that conclusion in 1957. Haldane pointed out that

[36]The qualification "basically" is only necessary because additive multiple locus models (that is, those with no epistasis) with complete linkage are also "beanbag" models in Mayr's sense. Formally, these are the same as single locus models with a higher number of alleles.

> this was what he had predicted in 1924 when it was not taken seriously. "If biologists had a little more respect for algebra and arithmetic", Haldane (p. 40) noted, "they would have accepted the existence of such intense selection thirty years before they actually did so";
>
> (iv) his own methods had first provided a technique for quantifying the type of evolutionary morphological change that Mayr found important.

The list goes on, each example detailing how, without mathematical analysis, verbal claims about evolutionary processes could not ever be directly appraised.

These examples were intended to pave the way for the fundamental point: "In the consideration of evolution, a mathematical theory may be regarded as a kind of scaffolding within which a reasonably secure theory expressible in words may be built up. I have given examples to show that without such a scaffolding verbal arguments are insecure" (p. 42). Moreover, mathematical analysis allows the exploration of different possible explanations of the same observed phenomenon. In the case of polymorphism, for instance, Fisher [1922] had shown how heterosis could maintain polymorphism, and this came to be generally accepted. But Haldane and Jayakar [1963] had shown that selection in fluctuating directions also sufficed to explain polymorphism. Though he did not believe that this mechanism was more frequent than Fisher's, he suggested that, had it been offered first, it might very well have commanded the greater allegiance. "The best way to avoid such contingencies", he argued, "is to investigate mathematically the consequences following from a number of hypotheses which may seem rather farfetched and, if they would lead to observed results, looking in nature or the laboratory for their truth or falsehood" (p. 44).

Finally, Haldane went on a counter-attack. He accepted that Mayr had made sympatric speciation far less probable than before. But he showed that Mayr's verbal arguments lacked precision and were sometimes so apparently contradictory that they could not even be formalized by those who favored precise reasoning. He accused Mayr of "a considerable ignorance of the earlier literature of beanbag genetics" (p. 48). For instance, Mayr had claimed that "the classical theory of genetics took it for granted that superior mutations would be incorporated into the genotype of the species while the inferior ones would be eliminated" [1963, 215]. Haldane pointed out that Fisher had shown the exact contrary, that heterosis could preserve polymorphism, in 1922 and that he, Haldane, had extended that result in 1926. He pointed out that Fisher [1918] had considered epistasis systematically and wondered why Mayr thought that the mathematical geneticists ignored such interactions. Lastly, he went on to provide a scathing assessment of Mayr's alternative framework:

> Mayr devotes a good deal of space to such notions as 'genetic cohesion', 'the coadapted harmony of the gene pool,' and so on. These apparently became explicable 'once the genetics of integrated gene complexes had replaced the old beanbag genetics'. So far as I can see, Mayr attempts

> to describe this replacement in his chapter ... on the unity of the genotype. ... This chapter contains a large number of enthusiastic statements about the biological advantages of large populations which, in my opinion, are unproved and not very probable. The plain fact is that small human isolates, whether derived from one 'race', like the Hutterites, or two, like the Pitcairn Islanders, can be quite successful. I have no doubt that some of the statements in Mayr's chapter ... are true. If so, they can be proved by the methods of beanbag genetics, though the needed mathematics will be exceedingly stiff. Fisher and Wright have both gone further than Mayr believes toward proving some of them. The genetic structure of a species depends largely on local selective intensities, on the one hand, and migration between different areas, on the other. ... The 'success' of a species can be judged both from its present geographical distribution and numerical frequency and from its assumed capacity for surviving environmental changes and for further evolution. I do not think that in any species we have enough knowledge to say whether it would be benefited by more or less 'cohesion' or gene flow from one area to another. We certainly have not such knowledge for our own species. ... Sewall Wright has been the main mathematical worker in this field, and I do not think Mayr has followed his arguments. (pp. 48–49)

His conclusion was: "In my opinion, beanbag genetics, so far from being obsolete, has hardly begun its triumphant career" (p. 49).

Among most evolutionary biologists, Haldane's defense settled the question for the next generation. With one notable exception, that of Ewens [1993], Mayr has received little support from scientists. Meanwhile, he opened a second front, now rebottling himself as a historian and philosopher of biology. From a sociological point of view, the transformation was successful but that issue is beyond the scope of this paper. In the late 1970s, Mayr, with Provine's help, organized a conference on the history of the evolutionary synthesis. It brought together historians and all the living scientists whom Mayr considered to have been important in the synthesis. In spite of Provine's involvement — and Provine was already working on Wright's biography — Wright was not invited.[37] At that conference, Lewontin [1980] delivered an uncompromising defense of mathematical population genetics, of the work of Fisher, Wright, and Kimura, but ignored Haldane. Carson [1990] underscored the significance of Haldane's *Causes of Evolution*, as noted before. Otherwise it was entirely Mayr's event. The naturalist account metamorphosed into its mature form with the synthesis allegedly occurring between 1936 and 1947. Mayr conveniently ignored the fact that in 1955 he had explicitly admitted that the naturalists had not yet successfully incorporated the necessary genetics into their framework. By now, he claimed that they had — with some help from others (but not mathematical population geneticists) — achieved the synthesis by 1947.

[37] Will Provine (personal communication).

Intentional or not, it was a remarkable exercise in self-service.

How credible is Waddington and Mayr's critique of theoretical population genetics? The evidence, on the whole, is decisively against them. Waddington made three independent claims: (i) that the mathematical work of Fisher, Haldane, Wright, and others had contributed "very few qualitatively new ideas"; (ii) that this work had not led to the discovery of any new theorems or conceptual relationships; and (iii) that it had not led to "any noteworthy quantitative statements about evolution". Mayr implicitly endorsed all of these claims. Explicitly, he emphasized the third one and added two new criticisms: (i) that the definition of fitness in these models ignored factors such as variable dominance, epistasis, temporal dependence of selection processes, *etc.*; and (ii) that the dynamics described by these models made a variety of unjustified simplifying assumptions.

Haldane's remarks from 1953 already provide ample response to Waddington's first claim. In particular, without Fisher's and Wright's pioneering mathematical analyses, it is hard to see how the idea of drift could have entered evolutionary theory. His second claim fares even worse. There is at least one conceptual relationship in evolutionary theory which has been regarded as interesting enough to have a cottage industry built around it [Crow, 1990a]. This is Fisher's so-called "Fundamental Theorem of Natural Selection". Though neither Wright nor Haldane brought it up in their responses to Mayr and Waddington, Fisher undoubtedly would have. But this theorem is not the only conceptual relationship to emerge from the mathematical work. The Hardy-Weinberg rule, the simplest and perhaps the most fundamental relation of population genetics has its origin in mathematics. Though Punnett suspected that, without selection (in a one-locus two-allele model with random mating and discrete generations), dominants would not gradually replace recessives (as was then commonly believed), he was unable to prove it without the mathematical help of Hardy [Punnett, 1950]. As Haldane noted in 1964, all such intuitions require mathematical exploration for their correctness to be judged.

Waddington's third claim was unjustified even in the early 1950s, as Section 3 of this paper amply documents. Quantitative work by Norton, starting around 1910 had shown how rapidly even small selective differences could establish a dominant trait, and how ineffective selection is for a recessive trait. Haldane's "Mathematical Theory" extended this work to more complicated models and systematically connected the models to field data. Haldane's examples in his defense of beanbag genetics underscores this point. What is less understandable is Mayr's emphasis on this criticism. Mayr's first new criticism, about the definition of fitness, was simply an error, as Wright pointed out. His second, about the simplifying assumptions made in the mathematical population geneticists' models, is correct. But, as is evident form Haldane's list of achievements, what is striking about mathematical population genetics is the extent to which it could provide useful results in spite of these assumptions.

However, Mayr's positive thesis about the achievements of the naturalists has more merit. There should be little doubt that the field and experimental work

of the Chetverikov school, of Dobzhansky, and of Mayr himself were important contributions to evolutionary biology. But, the question that Mayr never answered is why this work should not be regarded as, another naturalist, Sheppard, had regarded it in 1954, as providing experimental data testing the models of by mathematical population genetics.

What has so far been said recovers a central role for mathematical population genetics in the history of evolutionary theory. It remains to discuss Haldane's place in it relative to Fisher and Wright. But this will require some philosophical choices about the nature and role of theories. If it is accepted that global ("unifying") theories constitute the benchmark for what constitutes good theorizing, then it is arguable that Haldane was not as important as Wright and Fisher or, later, Kimura. But there are at least two reasons to question such a benchmark in general, and especially in the context of evolutionary theory:

(i) much of that theoretical enterprise consists of the construction of models to explore and test different proposed mechanisms of change: different models of hereditary transmission of genes (polyploidy, haplodiploidy, so on), relations between genes and traits, population structures, and modes of selection at different intensities. Here grand theory recedes into the background. What replaces them is, *first*, the strategy of model construction articulated first — and with full clarity — by Haldane in 1924 at the beginning of the "Mathematical Theory"; and *second*, to a surprising extent, many of the mechanisms that he first modeled (certation, kin selection, *etc*). Theoretical population genetics in practice proceeds largely by Haldane's methods, so much so, that his role in their origin has disappeared in a remote history that everyone has internalized to such an extent that any explicit mention is occasion for surprise. (The last point is similar to what Haldane [1953] said in his response to Waddington);

(ii) while universal theories may have many appeals, it is an independent question whether they are correct. No one knows whether the alternatives proposed by Fisher, Kimura, or Wright are correct in the sense that evolutionary history on Earth is best accounted for by any of their single preponderant mechanisms. Haldane did not think so and it may well turn out that this was a wise choice, that evolution has proceeded by a plurality of mechanisms, different ones in different ecological and genetical contexts, and the idea of a univocal theory is a chimera best cast aside as an inspired dream of a different and more heroic age. At the beginning of the twenty-first century, as modes and models of selection proliferate, it appears that Haldane, rather than Fisher or Wright was *methodologically* wise.

In 1990, Princeton University Press reissued *The Causes of Evolution* with annotations and a long "afterword" by E. G. Leigh. *Causes*, Leigh [1990] argued, "was eclipsed somewhat by the work of Fisher and Wright, apparently because Haldane refused to found a system" (p. xxii). Nevertheless, this was exactly

what Leigh found insightful and "light-hearted" (p. xx). Though Leigh admitted that "we need a system of sorts to make any sense at all of the real world", his ultimate conclusion was a ringing endorsement of Haldane over Fisher and Wright: "systems can blind us both to what has no place in them and to what conflicts with them" (p. xxiii). For Leigh, Haldane's major insight was his willingness to admit multiple levels of selection: the questions that can be asked in such models "are indeed the appropriate next stage in learning how adaptation can evolve" (pp. xxiii–xiv). In effect, Leigh underscores the points made in the last paragraph while also placing Haldane's work centrally within contemporary evolutionary debates.

ACKNOWLEDGMENTS

Thanks are due to James F. Crow, Warren Ewens, Raphael Falk, Steve Frank, M. J. S. Hodge, Mark Kirkpatrick, Steve Lanier, Richard C. Lewontin, Mohan Matthen, the late John Maynard Smith, the late Ernst Mayr, Anya Plutynski, William B. Provine, J. R. G. Turner, and William C. Wimsatt for many illuminating discussions and comments on earlier drafts of this paper which has had a long gestation. Much of the work on this paper was done while I was a Fellow at the Wissenschaftskolleg zu Berlin — I thank it for its support.

BIBLIOGRAPHY

[Adams, 1968] M. Adams. The founding of population genetics: Contributions of the Chetverikov School 1924–1934. *Journal of the History of Biology*, 1: 23–39, 1968.

[Adams, 1980] M. B. Adams. Sergei Chetverikov, the Kol'tsov Insitute, and the evolutionary synthesis. In E. Mayr and W. B. Provine (ed.), *The Evolutionary Synthesis*. Cambridge, MA: Harvard University Press, pages 242–278, 1980.

[Barrett, 1895–1902] C. G. Barrett. *The Lepiidoptera of the British Isles*. London: Reeve & Co., 1895–1902.

[Bateson, 1922] W. Bateson. Evolutionary faith and modern doubts. *Nature*, 109: 553–556.

[Belloc, 1920a] H. Belloc. A few words with Mr. Wells. *Dublin Review*, 166: 182–202, 1920a.

[Belloc, 1920b] H. Belloc,. Mr. Wells' 'Outline of History'. *London Mercury*, 3: 43–62, 1920b.

[Bowler, 1989] P. J. Bowler. *The Mendelian Revolution: the Emergence of Hereditarian Concepts in Science and Society*. Baltimore: Johns Hopkins University Press, 1989.

[Carson, 1980] H. L. Carson. Cytogenetics and the neo-Darwinian synthesis. In E. Mayr and W. Provine (eds.), *The Evolutionary Synthesis*. Cambridge, MA: Harvard University Press, pages 86–95, 1980.

[Crow, 1990a] J. F. Crow. R. A. Fisher, a centenniel view. *Genetics*, 124: 207–211, 1990.

[Crow, 1990b] J. F. Crow. Sewall Wright's place in Twentieth-Century biology. *Journal of the History of Biology*, 23: 57–89, 1990.

[Crow, 1992] J. F. Crow. Centenniel: J. B. S. Haldane, 1892–1964. *Genetics*, 130: 1–6, 1992.

[Edwards, 1977] A. W. F. Edwards. *Foundations of Mathematical Genetics*. Cambridge, UK: Cambridge University Press, 1977.

[Elton, 1927] C. S. Elton. *Animal Ecology*. London: Sidgwick and Jackson, 1927.

[Falk, 2001] R. Falk. The rise and fall of dominance. *Biology and Philosophy*, 16: 285–323, 2001.

[Fish, 1914] H. D. Fish. On the progressive increase of homozygosis in brother-sister matings". *American Naturalist*, 48: 759–761, 1914.

[Fisher, 1918] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52: 399–433, 1918.

[Fisher, 1922] R. A. Fisher. On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, 42: 321–341, 1922.

[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930.

[Fisher, 1934] R. A. Fisher. Professor Wright on the theory of dominance. *American Naturalist*, 68: 370–374, 1934.

[Fisher, 1935] R. A. Fisher. Dominance in poultry. *Philosophical Transactions of the Royal Society (London) B*, 225: 197–226, 1935.

[Fisher, 1938] R. A. Fisher. Dominance in poultry. Feathered feet, rose comb, internal pigment and pile. *Proceedings of the Royal Society (London) B*, 125: 25–48, 1938.

[Hagedoorn and Hagedoorn-Vorstheuvel la Brand, 1921] A. L. Hagedoorn and A. C. Hagedoorn-Vorstheuvel la Brand. *The Relative Value of the Processes Causing Evolution*. Hague: Martinus-Nijhoff, 1921.

[Haldane, 1919] J. B. S. Haldane. The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics*, 8: 299–309, 1919.

[Haldane, 1924a] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part I. *Transactions of the Cambridge Philosophical Society*, 23: 19–41, 1924.

[Haldane, 1924b] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part II. The influence of partial self-fertilization, inbreeding, assortative mating, and selective fertilization on the composition of Mendelian populations, and on natural selection. *Proceedings of the Cambridge Philosophical Society*,1: 158–163, 1924.

[Haldane, 1926] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part III. *Proceedings of the Cambridge Philosophical Society*, 23: 363–372, 1926.

[Haldane, 1927a] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part IV. *Proceedings of the Cambridge Philosophical Society*, 23: 607–615, 1927.

[Haldane, 1927b] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society*, 23: 838–844, 1927.

[Haldane, 1930] J. B. S. Haldane. A mathematical theory of natural and articficial selection. Part VI. Isolation. *Proceedings of the Cambridge Philosophical Society*, 26: 220–230, 1930.

[Haldane, 1931a] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part VII. Selection intensity as a function of mortality rate. *Proceedings of the Cambridge Philosophical Society*, 27: 131–136, 1931.

[Haldane, 1931b] J. B. S. Haldane. A mathematical theory of natural selection. Part VIII. Metastable populations. *Proceedings of the Cambridge Philosophical Society*, 27: 137–142, 1931.

[Haldane, 1932a] J. B. S. Haldane. *The Causes of Evolution*. London: Harper and Brothers, 1932.

[Abraham, 1932b] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part IX. Rapid selection. *Proceedings of the Cambridge Philosophical Society*, 28: 244–248, 1932.

[Haldane, 1933] J. B. S. Haldane. The part played by recurrent mutation in evolution. *American Naturalist*, 67: 5–19, 1933.

[Haldane, 1934] J. B. S. Haldane. A mathematical theory of natural and artificial selection. Part X. Some theorems on artificial selection. *Genetics*, 19: 412–429, 1934.

[Haldane, 1935] J. B. S. Haldane. The rate of spontaneous mutation of a human gene. *Journal of Genetics*, 31: 317–326.

[Haldane, 1937] J. B. S. Haldane. The effect of variation on fitness. *American Naturalist*, 71: 337—349, 1937.

[Haldane, 1938] J. B. S. Haldane. Forty years of genetics. In J. Needham and W. Pagel (eds.), *Background to Modern Science*. Cambridge, UK: Cambridge University Press, pages 223–243, 1938.

[Haldane, 1939] J. B. S. Haldane. The theory of the evolution of dominance. *Journal of Genetics*, 37: 365–374, 1939.

[Haldane, 1953] J. B. S. Haldane. Foreword. *Symposia of the Society for Experimental Biology*, 7: ix–xix, 1953.

[Haldane, 1956] J. B. S. Haldane. The theory of selection for melanism in Lepidoptera. *Proceedings of the Royal Society (London) B*, 145: 303–304, 1956.

[Haldane, 1957] J. B. S. Haldane. The cost of natural selection. *Journal of Genetics*, 55: 51–524, 1957.
[Hardy, 1908] G. H. Hardy. Mendelian proportions in a mixed population. *Science*, 28: 49–50, 1908.
[Harrod, 1951] R. F. Harrod. *The Life of Maynard Keynes*. London: Macmillan, 1951.
[Hogben, 1932] L. Hogben. Evolution re-stated. *Eugenics Review*, 24: 222–225, 1932.
[Huxley, 1922] J. S. Huxley. Kentucky bans evolution. *The Nation & the Athenaeum*, 31: 68–69, 1922.
[Huxley, 1942] J. S. Huxley.. *Evolution: The Modern Synthesis*. London: George Allen & Unwin, 1942.
[Jennings, 1912] H. S. Jennings. Production of pure homozygotic organisms from heterozygotes by self-fertilization. *American Naturalist*, 46: 487–491, 1912.
[Jennings, 1914] H. S. Jennings. Formulae for the results of inbreeding. *American Naturalist*, 48: 693–696, 1914.
[Jennings, 1916] H. S. Jennings. The numerical results of diverse systems of breeding. *Genetics*, 1: 53–89, 1916.
[Jennings, 1917] H. S. Jennings. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. *Genetics*, 2: 97–154, 1917.
[Kacser and Burns, 1981] H. Kacser and J. A. Burns. The molecular basis of dominance. *Genetics*, 97: 639–666, 1981.
[Keith, 1922a] A. Keith. Is Darwinism at the dusk or the dawn. *Nineteenth Century*, 92: 173–182, 1922.
[Keith, 1922b] A. Keith. Why I am a Darwinist. *Rationalist Annual*, 1922: 11–14, 1922.
[Kettlewell, 1973] B. Kettlewell. *The Evolution of Melanism: The Study of a Recurring Necessity*. Oxford: Clarendon Press, 1973.
[Kettlewell, 1956] H. B. D. Kettlewell. A résumé of investigations on the evolution of melanism in the Lepidoptera. *Proceedings of the Royal Society (London) B*, 145: 297–303, 1956.
[Kimura, 1956] M. Kimura. A model of a genetic system which leads to closer linkage by natural selection. *Evolution*, 10: 278–287, 1956.
[Kimura, 1983] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press, 1983.
[Lewontin and Kojima, 1960] R. C. Lewontin and K. I. Kojima. The evolutionary dynamics of complex polymorphisms. *Evolution*, 14: 458–472, 1960.
[Livingstone, 1922] B. E. Livingstone. The American Association at Toronto. *Nature*, 109: 285, 1922.
[Lotka, 1922] A. J. Lotka. The stability of the normal age distribution. *Proceedings of the National Academy of Sciences (USA)*, 8: 339–345, 1922.
[Lunn and Haldane, 1935] A. Lunn and J. B. S. Haldane. *Science and the Supernatural*. New York: Sheed & Ward, Inc., 1935.
[Majerus, 1998] M. E. N. Majerus. *Melanism: Evolution in Action*. Oxford: Oxford University Press, 1998.
[Maynard Smith, 1992] J. Maynard Smith. J. B. S. Haldane. In S. Sarkar (ed.), *The Founders of Evolutionary Genetics*. Dordrecht: Reidel, pages 37 -52, 1992.
[Mayr, 1942] E. Mayr. *Systematics and the Origin of Species*. New York: Columbia University Press, 1942
[Mayr, 1980] E. Mayr. Prologue: Some thoughts on the history of the evolutionary synthesis. In E. Mayr and W. B. Provine (eds.), *The Evolutionary Synthesis: Perspectives on the Unification of Biology*. Cambridge, MA: Harvard University Press, pages 1–48, 1980.
[Mayr, 1982] E. Mayr. *The Growth of Biological Thought*. Cambridge, MA: Harvard University Press, 1982.
[Mayr, 1992] E. Mayr. Haldane's *Causes of Evolution* 60 years later. *Quarterly Review of Biology*, 67: 175–186, 1992.
[Mayr and Provine, 1980] E. Mayr and W. B. Provine (eds.). *The Evolutionary Synthesis*. Cambridge, MA: Harvard University Press, 1980.
[McOuat and Winsor, 1995] G. McOuat and M. P. Winsor. J. B. S. Haldane's Darwinism in its religious context. *British Journal for the History of Science*, 28: 227–231, 1995.
[Morgan *et al.*, 1925] T. H. Morgan, C. Bridges, and A. H. Sturtevant. The genetics of Drosophila. *Bibliographia Genetica*, 2: 1-262, 1925.

[Muller, 1932] H. J. Muller. Further studies in the nature and causes of gene mutations. In D. F. Jones (ed.), *Proceedings of the Sixth International Congress of Genetics*. 1. Menasha: Brooklyn Botanic Garden, pages 213–255, 1932.
[Norton, 1928] H. T. J. Norton. Natural selection and Mendelian variation. *Proceedings of the London Mathematical Society*, 28: 1–45, 1928.
[Pearl, 1913] R. Pearl. A contribution towards an analysis of the problem of inbreeding. *American Naturalist*, 47: 577–614, 1913.
[Pearl, 1914a] R. Pearl. On a general formula for the constitution of the nth generation of a Mendelian population in which all matings are of brother X sister. *American Naturalist*, 48: 491–494, 1914.
[Pearl, 1914b] R. Pearl. On the results of inbreeding a Mendelian population; a correction and extension of previous conclusions. *American Naturalist*, 48: 57–62, 1914.
[Pearson, 1904] K. Pearson. Mathematical contributions to the theory of evolution. XII. On a generalized theory of alternative inheritance, with special reference to Mendel's laws. *Philosophical Transactions of the Royal Society (London) A*, 203: 53–86, 1904.
[Pearson, 1910] K. Pearson. Darwinism, biometry and some recent biology. I. *Biometrika*, 7: 368–385, 1910.
[Plunkett, 1932] C. R. Plunkett. Temperature as a tool of research in phenogenetics: Methods and results. In D. F. Jones (ed.), *Proceedings of the Sixth International Congress of Genetics*. 2. Menasha: Brooklyn Botanic Garden, pages 158–160, 1932.
[Plunkett, 1933] C. R. Plunkett. A contribution to the theory of dominance. *American Naturalist*, 67: 84–85, 1933.
[Provine, 1971] W. B. Provine.. *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press, 1971.
[Provine, 1985] W. B. Provine. The R. A. Fisher-Sewall Wright controversy. *Oxford Surveys in Evolutionary Biology*, 2, 1985.
[Provine, 1986] W. B. Provine. *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press, 1986.
[Punnett, 1915] R. C. Punnett. *Mimicry in Butterflies*. Cambridge: Cambridge University Press, 1915.
[Robbins, 1917] R. B. Robbins. Some applications of mathematics to breeding problems. *Genetics*, 2: 489–504, 1917.
[Robbins, 1918a] R. B. Robbins. Applications of mathematics to breeding problems II. *Genetics*, 3: 73–92, 1918.
[Robbins, 1918b] R. B. Robbins. Random mating with the exception of sister by brother mating. *Genetics*, 3: 390–396, 1918.
[Robbins, 1918c] R. B. Robbins. Some applications of mathematics to breeding problems III. *Genetics*, 3: 375–389, 1918.
[Robinson, 1922] J. H. Robinson. Is Darwinism dead? *Harper's Magazine*, 145: 68–74, 1922.
[Romanes, 1896] G. J. Romanes. *Life and Letters*. London: Longmans, Green, 1896.
[Sarkar, 1992a] S. Sarkar. "Haldane as Biochemist: the Cambridge Decade, 1923 -1932." In Sarkar, S. Ed. *The Founders of Evolutionary Genetics*. Dordrecht: Kluwer, pp. 53 -81, 1992.
[Sarkar, 1992b] S. Sarkar (ed.). *The Founders of Evolutionary Genetics*. Dordrecht: Kluwer, 1992.
[Sarkar, 1978] S. Sarkar. *Genetics and Reductionism*. New York: Cambridge University Press, 1998.
[Sarkar, 2004] S. Sarkar. Evolutionary theory in the 1920s: The nature of the synthesis. *Philosophy of Science*, 71: 1215–1226, 2004.
[Shaw, 1921] B. Shaw. *Back to Methuselah: A Metabiological Pentateuch*. London: Constable, 1921.
[Trow, 1913] A. H. Trow. Forms of reduplication–primary and secondary. *Journal of Genetics*, 2: 313–324, 1913.
[Trow, 1916] A. H. Trow. A criticism of the hypothesis of *linkage* and *crossing over*. *Journal of Genetics*, 5: 281–297, 1916
[Vavilov, 1932] N. I. Vavilov. The process of evolution in cultivated plants. In D. F. Jones, (ed.), *Proceedigs of the Sixth International Congress of Genetics*. I. Menasha, WI: Brooklyn Botanic Garden, pages 331–342, 1932.
[Warren, 1917] H. C. Warren. Numerical effects of natural selection acting upon Mendelian characters. *Genetics*, 2: 305–312, 1917.

[Weinberg, 1908] W. Weinberg. Über den Nachweis der Vererbung beim Menschen. *Jahreshelft Ver. vaterl. Naturf. Württemberg*, 64: 369–382, 1908.
[Weinberg, 1909a] W. Weinberg. Über Vererbungsgesetze beim Menschen. 1. Allgemeiner Teil. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre*, 1: 377–392, 440–460, 1909.
[Weinberg, 1909b] W. Weinberg. Über Vererbungsgesetze beim Menschen. 2. Spezieller Teil. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre*, 2: 276–330, 1909.
[Weinberg, 1910] W. Weinberg.. Weitere Beiträge zur Theorie der Vererbung. *Archiv für Rassen- und Gesellschafts-Biologie*, 7: 35–49, 169–173, 1910.
[Wells, 1923] H. G. Wells. *Men Like Gods*. London: Cassell, 1923.
[Wentworth and Remick, 1916] E. N. Wentworth and B. L. Remick. Some breeding properties of the generalized Mendelian population. *Genetics*, 1: 608–616, 1916.
[Willis, 1922] J. C. Willis. *Age and Area: A Study of Geographical Distribution and Origin of Species*. Cambridge, UK: Cambridge University Press, 1922.
[Wimsatt, 1992] W. C. Wimsatt. Golden generalities and co-opted anomalies: Haldane vs. Muller and the Drosophila group on the theory and practice of linkage mapping. In S. Sarkar (ed.), *The Founders of Evolutionary Genetics: A Centenary Reappraisal*. Dordrecht: Kluwer, pages 107–166, 1992.
[Wright, 1921a] S. Wright. Systems of mating. I. The biometric relations between parent and offspring. *Genetics*, 6: 111–123, 1921.
[Wright, 1921b] S. Wright. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics*, 6: 124–143, 1921.
[Wright, 1921c] S. Wright. Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics*, 6: 144–161, 1921.
[Wright, 1921d] S. Wright. Systems of mating. IV. The effects of selection. *Genetics*, 6: 162–166, 1921.
[Wright, 1929a] S. Wright. The evolution of dominance. *American Naturalist*, 63: 556–561, 1929.
[Wright, 1929b] S. Wright. Fisher's theory of dominance. *American Naturalist*, 63: 274–279, 1929.
[Wright, 1931] S. Wright. Evolution in Mendelian populations. *Genetics*, 16: 97–159, 1931.
[Wright, 1932] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In D. F. Jones (ed.), *Proceedings of the Sixth International Congress of Genetics*. I. Monisha, WI: Brooklyn Botanic Garden, pages 356–366, 1932.
[Wright, 1934a] S. Wright. Physiological and evolutionary theories of dominance. *American Naturalist*, 68: 25–53, 1934.
[Wright, 1934b] S. Wright. Professor Fisher on the theory of dominance. *American Naturalist*, 68: 562–565, 1934b.
[Wright, 1960] S. Wright. Genetics and twentieth century Darwinism: A review and discussion. *American Journal of Human Genetics*, 12: 365–372, 1960.
[Wright, 1968] S. Wright. Contributions to genetics. In K. Dronamraju, (ed.), *Haldane and Modern Biology*. Baltimore: Johns Hopkins Press, pages 1–12, 1968.
[Yule, 1902] G. U. Yule. Mendel's laws and their probable relations to intra-racial heredity. *New Phytologist*, 1(9): 193–207, 222–238, 1902.
[Yule, 1906] G. U. Yule. On the theory of inheritance of quantitative compound characters on the basis of Mendel's laws — A preliminary note. In W. Wilks (ed.), *Report of the Third International Conference 1906 on Genetics*. London: Royal Horticultural Society, pages 140–142, 1906.

SEWALL WRIGHT

James F. Crow

Sewall Wright was one of the great trio, R. A. Fisher, J. B. S. Haldane, and himself, who, starting about the time of World War I, founded the mathematical theory of microevolution and population genetics. Until the end of World War II, these three almost totally dominated the field. Wright's three most important contributions, all mathematical in nature, are: the theory of inbreeding and population structure, the shifting-balance theory of evolution, and the method of path analysis. But these are only part of a long and extremely productive scientific life. He was also a rarity among biologists in contributing to philosophy: his theory of organism and the mind-body problem.

After a brief biography, I discuss Wright's contributions to inbreeding theory, statistics, physiological genetics, population genetics, evolution, and philosophy. At the end I have added a short section on Sewall Wright the man — his character and personality.

Much of this material is taken from a more detailed article [Crow, 1992]. For more details of Wright's life, both personal and scientific, see Provine's book [Provine, 1986]. Wright has summarized in four volumes most of his own work as well as that of many others [Wright, 1968-78].

1 A BIOGRAPHY IN BRIEF

Sewall Wright was born in Melrose, Massachusetts on December 21, 1889. Shortly afterward, his father accepted a teaching position at Lombard College in Galesburg Illinois, where Sewall grew up. He was a precocious child, being able to read, write, and do arithmetic before starting school. He astonished his first grade teacher by knowing how to extract cube roots. At the teacher's request, he demonstrated this at the blackboard for the eighth grade class. The consequence, he said, was instant unpopularity with the other students. Thereafter he remained studiously quiet in class, rarely volunteering anything.

In due course he enrolled at Lombard College, where his father taught an astonishing variety of courses, including mathematics, surveying, astronomy, and English composition, in addition to his own subject, economics. Later he joined the Brookings Institute and wrote several books on economic policies. Sewall Wright had two brothers, Quincy who became widelly known as a scholar of international law and Theodore, equally well known in aeronautical engineering.

Many of Wright's college courses were taught by his father. Although most biologists regard Wright as a formidable mathematician, actually what he didn't

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

learn from his father was self-taught. At Lombard College Wright was fortunate to come under the stimulating influence of Dr. Wilhelmine Key — her husband was a descendant of Francis Scott Key. She steered Wright to the Encyclopedia Britannica. This was the great eleventh edition and the article on Mendelism was written by a leading British geneticist, R. C. Punnett. She also expanded his horizons and encouraged him to think of a career as a geneticist.

After his junior year, Wright spent a year (1909-1910) as a surveyor's assistant for a railroad line being built in the Standing Rock Reservation in South Dakota. His mathematical skills were soon evident and he did the calculations for the proper angles and alignment of the rails on turns. It was a time of great adventure in the spirit of the old west with cowboys, Indians, mule-skinners, and outlaws. In his later years he loved to talk about this and in his nineties he remembered an astonishing number of words in the Sioux dialect. During this year, he developed a lung infection, which necessitated his living in a caboose; he used the leisure to read mathematics, including a book on quaternions. The book still exists. It is clear that he read it with understanding and it is obvious exactly how far he got. Curiously, Haldane read the same book, and as far as I know neither of them ever made use of quaternions. Wright's lung infection caused him to have trouble getting life insurance, something that he found increasingly amusing as he lived well into his nineties.

After a year at the University of Illinois, he joined the laboratory of William Castle at Harvard. He received his PhD in 1915 and was immediately employed as animal husbandman at the United States Department of Agriculture (USDA) near Washington, D. C. In 1926 he moved to the University of Chicago where he stayed until he reached the mandatory retirement age of 65. He then moved to the University of Wisconsin where he continued to work for an additional 30 years. For more details, see [Provine, 1986].

In his retirement, Wright wrote the four volumes summarizing his life work [Wright, 1968-78]. Wright continued to read and write until bad eyesight forced him to give this up. He died prematurely in his ninety ninth year on March 3, 1988. I say prematurely because he was in robust health and looking forward to a centennial celebration, which his colleagues were already planning. His poor eyesight was the problem; on one of his walks he failed to see an icy spot and died as the result of a fall.

Wright's last paper [Wright, 1988] was a summary of his evolutionary views, placing it in the context of the work of others and responding to some criticisms in Provine's book [Provine, 1986]. The reprints came while he was in the hospital for the last time and he had explicit instructions as to where to send them.

2 INBREEDING

Every student of introductory genetics learns Wright's inbreeding coefficient and a simple algorithm for computing it [Wright, 1922a]. There is a story behind this. In the early years of the 20th century a number of people worked on the

consequences of various pedigrees involving mating of relatives and the consequent decrease of heterozygosity. One problem was repeated mating of brother and sister. The usual method was to grind out the detailed consequences for a few generations and hope to find a pattern. In later generations, the problem soon beomes unwieldy. One well-known geneticist got it wrong by not carrying the process through enough generations. When Wright was starting his graduate work at Harvard, his roommate, H. D. Fish, was working on this problem. Fish had covered not only his desk but much of the floor with calculations. Wright thought there must be a better way, and he found it.

He devised the now-familiar algorithm which uses a simple process of tracing paths of ancestry from one parent to a common ancestor and back to the other parent. The whole process is easily learned and routinized, so it is readily adapted to computer use for complicated pedigrees, including tens of thousands of records for an entire breed of livestock. The inbreeding coefficient is a measure of the proportion by which the heterozygosity of an individual is reduced by inbreeding. Although Wright derived his algorithm by using correlation coefficients, this derivation has been largely replaced by using the concept of "identity by descent (IBD)". Two homolgous alleles are IBD if they are both descended from an allele in a common ancestor, or one is descended from the other. This has been extensively used in animal and plant breeding and recently, with the coming of molecular chromosome markers and computers, the concept is now widely used in the study of human genetics.

Wright made two important applications of inbreeding theory while he was at the USDA. The first was experimental. He inherited a colony of guinea pigs that had been maintained, usually with close inbreeding, for many generations. He analyzed the effects of inbreeding on a variety of important traits, such as size, vigor, and fertility. The analysis was masterful and these papers, published in 1922, have stood the test of time; they can be read with profit today [Wright, 1922b].

The second application was to the study of breeds of livestock. Wright subdivided the amount of inbreeding into components such as local consanguineous mating and herd differences. He found that almost all the increased homozygosity was due to random differences between herds rather than non-random mating within a herd. He also found considerable genetic differentiation among herds. His analysis of the extensive records of Shorthorn cattle is classic.

Wright extended this work into what are called F-statistics. The inbreeding coefficient is generalized to include population structure, particularly of a hierarchical sort. These measures are widely used in the study of natural populations, including humans. Both the cattle studies and later theory are summarized by Wright in a 1951 paper [Wright, 1951].

These studies led Wright to what he regarded as his major accomplishment, the "shifting balance theory" of evolution. Wright noticed that most of the breed improvement arose by random differences among the herds, followed by exporting bulls from the best herds and thereby upgrading the whole breed. His evolutionary

view grew out of this; I shall return to this subject later.

3 STATISTICS AND PATH ANALYSIS

Wright had an interest in statistics from his earliest scientific studies. He was familiar with the work of Karl Pearson and soon became adept in statistical methods, especially in the use of the correlation coefficient. This proclivity lasted through his life; Wright preferred to attack almost any problem through correlations. An example, inbreeding, has already been mentioned. His derivation of the inbreeding coefficient by correlations seems awkward and indirect to many modern workers, but that was how he worked. In fact, he defended his procedure, since he liked the idea that inbreeding could be negative, if mates were less closely related than random pairs. Correlations can be negative; probabilities, of course, cannot.

Wright was ahead of the field in two early statistical studies. In his early study of guinea pig weights in 1917, he was able to subdivide product moments into between- and within-group components [Wright, 1917]. In effect, he had discovered what was later called the analysis of covariance, a subject invented independently and carried farther by R. A. Fisher. Another example: In a study of the extent of white-spotting in guinea pigs he found a transformation to linearize cumulative percentage data [Wright, 1926]. This later came to be known as probit analysis [Finney, 1971].

Wright's most important contribution to statistical methodology is his method of path analysis [Wright, 1921]. The algebra is that of partial regression analysis, but the approach is unusual and typical of Wright. He was not particularly interested in correlation and regression as descriptive or predictive. Rather, his main interest was interpretative; he was especially interested in causal analysis. He assumed that the causal pathways were known but the relative magnitude of different paths was to be determined. His procedure was to draw diagram with a sequence of paths. A path of influence was indicated by an arrow while unanalyzed correlations were indicated by two-headed arrows. This was a natural way to diagram pedigrees, but Wright extended it to other problems. Each step in a pathway was associated with a "path coefficient", a partial regression coefficient standardized by being measured in standard deviation units. A path coefficient is then a measure of the relative contribution of this particular step in the pathway. Wright devised simple rules from which it is easy, from the path diagram, to write the relevant equations, which can then be solved for the unknown variables. A virtue of the method is that it immediately shows whether there are sufficient measurements and equations to permit a unique solution.

In 1925 Wright published a monumental analysis of the production and prices of corn and hogs in the period from the Civil War to World War I [Wright, 1925]. This involved no less than 510 correlations. Wright did the calculations himself, using the primitive card-sorting and calculating equipment available at the time. It would be a quick calculation with modern computers, but at the time this required an enormous amount of time. Remarkably, he accounted for some 80

percent of the variance of hog production and prices by fluctuations in the corn crop and various correlations. In several ways this was ahead of its time, especially by introducing time lags, later used extensively by economists.

This work was done in the early 1920s, but, incredibly, Wright was not permitted to publish it. An animal husbandman had no business working on economic problems. The manuscript lay unpublished for several years. It required the intervention of Henry Wallace, later Vice President of the United States, to persuade his father, then Secretary of Agriculture, to allow the paper to be published. It has long been regarded as a landmark, especially by economists.

In the early years after Wright's discovery of the method, path analysis was very popular among animal breeders. This was largely due to the strong influence of Jay L. Lush, an animal breeder at Iowa State College, who was a great admirer of Wright and a great popularizer of his work. Path analysis is best adapted to analyzing data from natural populations rather than planned experiments. Later, as animal breeding became more regularized, other methods that were more adapted to computer analysis and to tests of statistical significance took over, so path analysis is now something of a rarity among animal breeders. Meanwhile, it has been taken up by social scientists and is now used by sociologists, economists, and philosophers.

4 PHYSIOLOGICAL GENETICS

For most of his working life, Wright studied guinea pigs. His major contributions were to gene action and interaction, what was then called physiological genetics. While at USDA and the University of Chicago, a large part of his working day was spent with his guinea pigs. He did the husbandry and record-keeping himself. His extensive records still exist and have recently been transferred from Wisconsin to the custody of William Provine, a historian of genetics, who wrote the definitive biography of Wright [Provine, 1986], including the work of this section. Wright's physiological interest comes as a surprise to most geneticists, since his inbreeding, population structure, and evolutionary work are much better known.

Wright's initial work on guinea pigs was an accident of timing. William E. Castle, his mentor at Harvard, worked on various rodents and assigned each incoming graduate student to a particular species. C. C. Little worked on mice, Castle himself studied rats, E. C. MacDowell studied rabbits, and John Detlefson, guinea pigs. Wright arrived just as Detlefson was leaving so naturally he began working on guinea pigs. He never willingly stopped. When he moved from Chicago to the University of Wisconsin he had to give up the guinea pig colony for lack of facilities. He regretted this, but I am sure that this was for the best, since he was then free to write. His four-volume treatise on evolution and population genetics, written in his eighties, might never have been completed if he had had his beloved guinea pigs to work with. As it was, Wright spent the first five years in Wisconsin completing analysis of his guinea pig records.

Soon after joining the USDA Wright wrote a series of articles in the Journal

of Heredity on coat color inheritance in mammals [Wright, 1917-1918]. These included rats, guinea pigs, mice, cattle, horses, swine, dogs, cats, and humans. This early analysis was remarkable in two ways. One was his finding similarities and presumed homologies among genes with similar effects in these species; many have since been confirmed. The second is his interpretation of these colors in terms of the latest knowledge of enzymes and pigment chemistry. As often, he was ahead of his time.

Wright's approach was always quantitative. An early paper analyzed the inheritance of size factors into components, based on correlation of various body parts. He was able to subdivide the variance of size into components of (1) general size, (2) limb-specific factors, (3) fore- and hindlimb factors, (4) upper- and lower-limb factors, and (5) special factors for each part [Wright, 1918]. This was the forerunner of several papers by various authors, who did similar analyses in a number of species.

Wright's deepest interest was in the interaction of genes affecting coat color in guinea pigs. He formulated the relationship in path diagrams assuming flux equilibrium and expressed the various chemical processes as differential equations. His papers showed his mastery of the latest concepts of enzyme chemistry. Wright's major analysis of this kind was presented in 1941 [Wright, 1941]. He had plans for a book on this subject. But 1941 was the year that biochemical mutants in Neurospora were discovered, and this opened up the opportunity for a much more direct attack on gene-enzyme relations. I think Wright realized that this represented a new direction in genetics, as indeed it did, for it led to the discovery of sex in bacteria and ultimately to molecular genetics and the nature of the gene. Wright continued to do guinea pig studies, with increasingly elaborate analyses of the action and interaction of the various genes. These studies were masterful, but their great detail made them difficult to read. In fact they attracted very little attention. The fashion had changed and Wright's kind of analysis was replaced by molecular biology and direct chemical analysis.

One point should be mentioned. Wright was often impressed by the frequent unpredictability of interactions. Often the combined effect of two genes was not at all like what would be expected from their individual effects. Such examples had a large effect on Wright's thinking, especially his shifting balance theory of evolution.

5 POPULATION GENETICS

As mentioned earlier, population genetics was founded and almost completely dominated for a number of years by Haldane, Fisher, and Wright. Their styles were quite distinctive. Haldane wrote a series of papers that systematically explored selection under a wide variety of circumstance. These were mainly single locus models and the mathematics was relatively simple, summarized in [Haldane, 1932]. Haldane is also responsible for the concept of "mutation load", the idea that the impact of mutation on the population is determined mainly by the mutation rate

and not by the effect of the individual mutations [Haldane, 1937]. A second idea is his showing that the amount of reproductive excess required to carry through an allele substitution is a function mainly of its initial frequency [Haldane, 1957]. Both of these have the Haldane touch of finding an unexpectedly simple relationship, but both have turned out to be too restricted for most modern applications.

Fisher's best-known contribution to animal breeding and evolution is his showing by the use of least-squares theory how selection, either natural or man-made, can improve a population despite complicated interactions among genes [Fisher, 1918]. The central idea in his view was his "Fundamental Theorem of Natural Selection", namely that the rate of change of mean fitness of a population at any instant is determined by the genic variance of fitness at that time. Genic variance is the variance of the additive component of the phenotype, as determined by least-squares. Fisher was also the first to introduce stochastic processes into genetics. His block-buster was his "The Genetical Theory of Natural Selection", regarded by some as the greatest book on evolution since Darwin [Fisher, 1926]. Whether this is true, there is no doubt as to the richness of the ideas; each rereading brings new insights. Fisher was often obscure, but nobody could match his elegance both in mathematics and in English prose. If you read the book, use the variorum edition, which has a very useful appendix that clears up a number of Fisher's obscurities [Fisher, 1926].

Wright was always interested in population structure. His inbreeding coefficient, discussed earlier, and its extension to hierarchical population structure are among his greatest accomplishments, now widely used in the study of population structure, including humans [Wright, 1951]. One of his aims was to write in general form an equation describing the change of gene frequency under the influence of mutation, migration, selection, and random processes. It later turned out that his equation, in differential form, was already in use by physicists — the Fokker-Planck equation applied to diffusion and other processes that involve deterministic and random elements [Wright, 1968-78].

None of the three pioneers was concerned with mathematical rigor. After World War II the field of population genetics became more mathematical. The French mathematician, Gustave Malécot, developed Fisher and Wright's ideas in a mathematically rigorous form [Malécot, 1948]. Motoo Kimura introduced the Kolmogorov partial differential equations to solve problems that had eluded Wright. Kimura's most important papers have been collected in a single volume [Kimura, 1994]. This work laid the foundation for the realization that a large part of DNA changes in evolution is driven by mutation and random processes and is relatively uninfluenced by selection [Kimura, 1983]. This then led to a "molecular clock" by which rates of evolutionary change can be calibrated.

Although most of Wright's work in population genetics was theoretical, he entered several collaborations with experimentalists, especially Th. Dobhansky. This was the beginning of a trend toward theory-driven experimentation in population genetics that still exists. Wright soon tired of this and, sometimes with difficulty, disentangled himself from these alliances, and went back to his theoretical studies

[Provine, 1986].

6 THE SHIFTING BALANCE THEORY

Wright regarded his evolution work as his most important contribution. The general idea goes back to 1920s and, although it underwent a number of refinements, the basic principle remained unchanged. In later writings, he called this the "Shifting Balance Theory". For a very full analysis, see [Wright, 1968-78].

The basic difficulty, as Wright saw it, is that natural selection cannot ordinarily change a sexual population from a state of high average fitness to one of higher fitness if this means passing through a stage with lesser fitness. The shifting balance theory is an attempt to get around this difficulty.

According to Wright, the theory came mainly from three observations.

(1) In his analysis of guinea pig hair pattern and coat colors Wright was impressed by the fact that genes in combination are often not what would be predicted from their individual effects. An example that he often mentioned was a rosette hair pattern, in which the combined effect of two genes was opposite to what would have been predicted from their individual effects.

(2) His observations of inbred lines of guinea pigs showed that inbreeding, in addition to bringing about decline in vigor and fitness, also led to differentiation among the inbred lines. The variability among the lines was much greater than that in the foundation stock from which the inbreds were derived.

(3) The third observation came from the history of domestic livestock, particularly shorthorn cattle. He saw that advances in the quality of the breed did not seem to arise from recurrent selection by breeders. Rather it sometimes happened that, for no apparent reason, a particular herd demonstrated superior qualities. Bulls exported from this herd transmitted these qualities to the whole breed so that the breed gradually attained the quality of the favored herd. Then the cycle could be repeated, as another herd drifted into having desirable qualities.

These observations are almost a summary of the shifting balance theory. Wright liked to use a peaks and valleys metaphor to describe the process as he saw it. In such a three dimensional graph the abscissas are usually allele frequencies and the ordinate is the mean fitness of a population with corresponding allele frequencies. The loci are treated as independent, making the abscissas orthogonal. The idea is easily generalized, if not easily visualized, to more than three dimensions.

Wright thought of the shifting balance process in three phases.

> *Phase 1.* He envisaged a large population broken up into partially isolated subpopulations (like herds of cattle). These subpopulations are small enough that allele frequencies within a population drift randomly. Somewhere among the subpopulations, one may happen to drift into a favorable combination of allele frequencies.
>
> *Phase 2.* Once a subpopulation comes into the domain of attraction of a higher peak, the gene frequencies will change such that the subpopu-

> lation population "climbs" the higher peak. Although phase 1 depends on random processes, phase 2 is simple selection and moves relatively rapidly.
>
> *Phase 3.* A subpopulation at a higher peak, being more fit, will increase in numbers relative to other subpopulations and therefore, perhaps because of crowding, export migrants. These migrants upgrade the subpopulations into which they have migrated, much as the Shorthorn bulls exported from a superior herd raised the breed quality. In this way the entire population is upgraded, and the process can start over.

Wright did a great deal of mathematics on phases 1 and 2, but not phase 3. However more recent work has shown that, at least in some cases, phase 3 can work as Wright envisioned [Crow *et al.*, 1990]. Repeated directional migration can be a strong force.

One criticism of Wright's theory is that it places quite stringent requirements on the population structure. In the first place there must be several subpopulations, the more of these, the better for the theory. The individual subpopulations must not lose the variability necessary for random gene-frequency drift, so some migration from other colonies is required. Yet, too much migration leads to loss of subdivision; the population becomes a single unit and the process fails. So a major criticism of Wright's theory is that the conditions required for it to work optimally will not occur very often.

A second objection is that phase 1 reduces the population fitness, since random changes are mainly maladaptive. It is not appealing, at least to some evolutionists, that a process intended to increase fitness can function only by at least temporarily reducing it.

A third criticism is that a more appropriate metaphor may not be a rugged landscape but rather an ocean in which the surface is constantly changing. The idea is that, given an ever-changing environment, it is very unlikely that a population will ever be in such a state that *no* allele frequency change can increase fitness.

Finally, mass selection, often aided by artificial insemination, has been strikingly successful in improving livestock performance. One wonders whether if Wright had observed modern improvement in milk production of dairy cattle by mass selection, rather than the process he observed in the history of Shorthorns, he would have developed the shifting-balance theory.

None of these criticisms deterred Wright. He did not regard his process as necessarily of frequent occurrence, but as a way in which some population somewhere can bypass the difficulty of changing from a lower fitness peak to a higher one when the intermediate stage is maladapted. It was his preferred way of generating evolutionary novelty.

Wright's theory has long been very popular among biologists, although less so among theoretical population geneticists. Recently, it has come into more and more criticism, e.g. [Coyne *et al.*, 2000], so I think it is fair to say that the

theory is less widely accepted than it was in Wright's lifetime. It is not clear at present whether what Wright regarded as his crowning achievement will be his lasting heritage. Whether or not this is true, his high place in biological history is assured. His contributions to the mathematical theory of population genetics and to statistics are here to stay, as is his analysis of inbreeding. As one writer [Kimura, 1983] said: "Who cannot have wished to invent something so simple and important as the inbreeding coefficient?"

7 PHILOSOPHY

In 1952, as President of the American Society of Naturalists, Wright surprised his audience, who had heard him repeatedly speak about evolution, by discussing philosophy [Wright, 1953]. Biologists don't often do this.

Wright had a long interest in the philosophy of organism and the mind-body problem. His solution was an unusual one. He refused to accept emergence, in particular the emergence of mind. He used the word "organism" very broadly. Thus an organism could be an individual plant or animal, a hive of bees, a human society, a species, an ecosystem, or the entire biota. It could also be the earth, the solar system, or the universe. Or in the other direction, molecules, electrons or protons. If he were writing today, he would include quarks. Since he wanted to avoid emergence, he emphasized the continuity between individual, organ, cell, virus, and gene — and, more broadly, between living and non-living.

With a hierarchical organization where there were no clear borders between the different levels, Wright could not imagine truly emergent properties, such as mind. There is no place, he said, at which one can say that mind exists only beyond this point, whether one is speaking of stages of development, or times in evolutionary history, or even from quark to the universe. Emergence of mind from no mind, Wright said, is "sheer magic". "If the human mind is not appear by magic, it must be a development from the mind of the egg and back of this, apparently, of the DNA molecules of the egg and sperm nuclei that constitute its heredity" [Wright, 1964]. Seeing no stopping point here, Wright went on to say: "Because of the hierarchic nature of biologic and physical entities it appears that my mind must be based somehow on the minds of my cells and these on those of the constituent molecules and so on down to elementary particles." It sounds a bit like Leibniz.

According to Wright, one mind sees another as matter, except as they compare experiences. Wright thus arrived at his view of "dual-aspect panpsychism"; mind is everywhere. "The only satisfactory solution of these dilemmas would seem to be that mind is universal, present not only in all organisms and in their cells but in molecules, atoms and elementary particles" [Wright, 1964]. Mind is universal and so is matter.

As to the influence of this view on the practice of science, he wrote: "Acceptance of this point of view requires relatively little change in the actual practice of science, especially as determinism has never been more than an ideal admittedly unrealizable in full because of the invariable errors of observation and in

many cases, practically irreducible probabilities like those in the fall of dice (or segregation and assortment of genes). The deterministic expressions do not lose their usefulness as approximations. What we are given is a tenable philosophy of science and along with this a desirable humility in the recognition that science is a limited venture, concerned with the external and statistical aspects of events and incapable of dealing with the unique creative aspect of each individual event" [Wright, 1964]. Similar ideas are in [Wright, 1978]. Wright freely admitted that his philosophy had no influence on his every day practice of science.

Biologists have greeted this work with disagreement, or more often, with indifference. Many biologists regard the mind-body problem as something to avoid, better left to philosophers. Others regard mind as something growing out of matter, requiring only a certain level of organizational complexity. Most, I think, regard the mind as composed of constituents that cannot think at all. Wright did, however, find agreement from some philosophers, notably his long time friend, Charles Hartshorne.

Wright's 1964 paper [Wright, 1964] was written in honor of Hartshorne.

8 WRIGHT, THE MAN

Those who knew Wright from his early days say that he was always somewhat shy and introverted. He had absolutely no small talk, and it was difficult to have a casual conversation with him. In contrast, if the subject was serious, he could and would go on at great length. The topic could be history — including his early days surveying; his ancestors (he enjoyed mentioning that he was a descendant of Samuel Sewall, of Salem witchcraft fame); politics, on which he had strong opinions (he detested Nixon); and of course any kind of science. His intellectual interests extended beyond the usual range. He had limitations in aesthetics, however. He sometimes remarked that he had no interest in poetry, something that disappointed his father.

Wright was an unusually kindly man, always willing to help. No doubt he hurt his own career by spending inordinate amounts of time helping others; usually this involved a complete reanalysis of the data. He took manuscript reviewing extremely seriously. On one occasion in which I was involved as an editor, he reanalyzed the data and reached a conclusion opposite to that of the author. The author simply changed the conclusion.

Although Wright was kindly to almost everyone, there were two exceptions. One was R. A. Fisher, with whom he disagreed about several evolutionary points. For some reason not apparent to outsiders, the difference became intensely personal, and they avoided each other. Wright also complained about Ernst Mayr, who he said did not understand his (Wright's) work. Usually, however, he could see good in almost everyone and every paper. Often he had insights deeper and better than the original author.

Wright routinely used a desktop mechanical calculator and eschewed the use of electronic computers. Finally the Genetics Department bought him an electronic

calculator, which at first he was reluctant to use. But eventually he began experimenting with it. Once I was visiting his ill wife in a hospital when Wright came in. She asked him if he had emptied the trash, taken out the garbage, and done a few other things, none of which he had done. She said: "Sewall, you are the world's worst husband." He replied, "I may be a bad husband, but I know what factorial 69 is." The significance is that this is the largest factorial that does not overflow his computer.

For a man of such accomplishment, Wright was excessively modest with a self-deprecating wit, although he did not hesitate to argue a point on which he had an opinion — like politics or the shifting balance theory. I have several times repeated my favorite Wright anecdote, and will relate it here once again. While writing his books he received a modest stipend from the National Science Foundation. During this time the Foundation decided to provide an inflationary adjustment to his pay. He was in his late 80s at the time. When I brought him this good news, he replied that, according to his careful calculations, his productivity was declining at exactly the same rate as the value of the dollar and he didn't deserve a raise. He never accepted it.

BIBLIOGRAPHY

[Coyne *et al.*, 2000] J. A. Coyne, N. H. Barton, and M. Turelli. Is Wright's shifting balance process important in evolution? *Evolution*, 54: 306–317, 2000.

[Crow, 1992] J. F. Crow. Sewall Wright's place in twentieth-century biology. In S. Sarkar (ed.), *The Foudations of Evoltionary Genetics*, pages 167-200, Kluwer Academic, Dordrecht, 1992.

[Crow *et al.*, 1990] J. F. Crow, W. R. Engels and C. Denniston. Phase three of Wright's shifting balance theory. *Evolution*, 44: 233–247, 1990.

[Finney, 1971] D. J. Finney. *Probit Analysis*. Cambridge Univ. Press, Cambridge, 1971.

[Fisher, 1918] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Royal Soc. Edinb.*, 52: 399–433, 1918.

[Fisher, 1926] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford Univ. Press, Oxford, 1926. Complete Variorum Edition, Oxford Univ. Press, Oxford, 1999.

[Haldane, 1932] J. B. S. Haldane. *The Causes of Evolution*. Longmans, London, 1932.

[Haldane, 1937] J. B. S. Haldane. The effect of variation on fitness. *Amer. Natur.*, 71: 337–349, 1937.

[Haldane, 1957] J. B. S. Haldane. The cost of natural selection. *J. Genet.*, 55: 511–524, 1957.

[Hill, 1996] W. G. Hill. Sewall Wright's 'Systems of mating'. *Genetics*, 143: 1499–1506, 1996.

[Kimura, 1983] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge Univ. Press, Cambridge, 1983.

[Kimura, 1994] M. Kimura. *Population Genetics, Molecular Evolution, and the Neutral Theory*. Univ. Chicago Press, Chicago, 1994.

[Malécot, 1948] G. Malécot. *Les mathématiques de l'hérédité*. Masson et Cie, Paris, 1948.

[Provine, 1986] W. B. Provine. *Sewall Wright and Evolutionary Biology*. Univ. Chicago Press, Chicago, 1986.

[Wright, 1917] S. Wright. The average correlation within subgroups of a population. *J. Wash. Acad. Sci.*, 7: 532–535, 1917.

[Wright, 1917-1918] S. Wright. Color inheritance in mammals. *J. Heredity*, 8: 224–235, 373–378, 426–430, 473–475, 476–480, 521–527, 561–564, 1917; 9: 33–38, 89–90, 139–144, 227–240, 1918.

[Wright, 1918] S. Wright. On the nature of size factors. *Genetics*, 3: 367–374, 1918.

[Wright, 1921] S. Wright. Correlation and causation. *J. Agric. Res.*, 20: 557–585, 1921.

[Wright, 1922a] S. Wright. Coefficients of inbreeding and relationship. *Amer. Natur.*, 56: 330–338, 1922.

[Wright, 1922b] S. Wright. The effects of inbreeding and crossbreeding on guinea pigs. *Bull. No. 1090, U. S. Dept. Agric.*, 1922.
[Wright, 1925] S. Wright. Corn and hog correlations. *Bull. No. 1300 U. S. Dept. Agric.*, 1925.
[Wright, 1926] S. Wright. A frequency curve adapted to variation in percentage occurrence. *J. Amer. Stat. Assoc.*, 21: 161–178, 1926.
[Wright, 1941] S. Wright. The physiology of the gene. *Physiol Rev.*, 21: 487–527, 1941.
[Wright, 1951] S. Wright. The genetic structure of populations. *Ann. Eugen.*, 15: 323-354, 1951.
[Wright, 1953] S. Wright. Gene and organism. *Amer. Natur.*, 87: 5–18, 1953.
[Wright, 1964] S. Wright. Biology and the philosophy of science. *Monist*, 48: 265–290, 1964.
[Wright, 1978] S. Wright. Panpsychism and science. In D. R. Griffen and J. E. Cobb (eds.), *Mind in Nature*, pages 79–88, Univ. Press of America, Washington, 1978.
[Wright, 1968-78] S. Wright. *Evolution and the Genetics of Populations.* I (1968), II (1969), III (1977), IV (1978), Univ. Chicago Press, Chicago.
[Wright, 1988] S. Wright. Surfaces of selective value revisited. *Amer. Natur.*, 131: 115–123, 1988.

This page intentionally left blank

MOTOO KIMURA

James F. Crow

Motoo Kimura's work can be divided into two parts. The first is development of theoretical population genetics, especially stochastic processes. He was the first to apply both Kolmogorov equations, forward and backward [Feller, 1950], to genetic problems. In this and other ways he extended the work of the founders, R. A. Fisher, J. B. S. Haldane, and Sewall Wright.

The second part of his life, starting in 1968, was devoted to the "neutral theory". From a study of molecular evolution he concluded that most such evolution is driven by mutation and random drift rather than natural selection. This was initially controversial and was widely criticized, but it has come to be recognized as a part of population genetics theory, with wide applicability.

I shall discuss these two phases of his life, but initially I give a short biography and finally a short account of Kimura, the man and his personality. He was my student, colleague, and friend.

A note on pronunciation: The correct pronunciation of the second syllable of Kimura's given name is with a long vowel; in Japanese this is stretched out. He first indicated this with a circumflex. Later, doubling the letter became popular, and he spelled the name Motoo. Inevitably, in the United States it was mispronounced Mo-TOO rather than Mo-TOE. Of course his friends soon learned better, but throughout his life Kimura was constantly encountering mispronunciation of his name.

1 A BRIEF BIOGRAPHY

Motoo Kimura was born in Okazaki, Japan, on November 13, 1924. His father was a metal worker, descending from a long line famous for large bells. He loved flowers and imparted this love to his son. When Motoo was quite young his father gave him a microscope and Motoo spent many happy hours with it. He particularly liked plants and aspired to be a botanist.

During his childhood, an epidemic of food poisoning affected his whole family, and one brother died. During his enforced absence from school, Motoo studied his math texts, particularly Euclidean geometry. When he returned to school, his teacher and classmates were astonished that he had done all the problems. The teacher recognized his unusual ability and encouraged him to become a mathematician, but his stronger interest was botany. At the time there seemed to be no connection between mathematics and systematic botany — how different it is today.

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

Having passed the difficult entrance examinations, Kimura was admitted to the prestigious National High School in Nagoya. There he was fortunate in having as botany teacher, M. Kumazawa, a distinguished plant morphologist. Recognizing Kimura's ability, Kumazawa encouraged him to work in his lab. Cytology was of central interest in Japan at the time and Kimura soon began a serious study of lily chromosomes. Kumazawa also taught a course in biometry and for the first time Kimura realized, to his pleasure, that there was a place for mathematics in biology. He also enjoyed physics and his hero was the famous Japanese physicist, Hideki Yukawa. Despite his work in mathematical genetics, Kimura never lost his love of flowers. Much later, he used the royalties from a book we wrote together [Crow and Kimura, 1970] to build a greenhouse and grow orchids.

While Kimura was in high school, Japan was deeply involved in World War II. After finishing a war-shortened curriculum in 1944, Kimura was admitted to Kyoto Imperial University. The best-known genetics faculty member was Hitoshi Kihara, famous throughout the world for his studies on the evolution and cytogenetics of wheat. One would expect that Kimura would have joined Kihara's group in the Faculty of Agriculture and studied cytology. At this time, however, students in the Faculty of Science were exempt from military duty, so Kimura, at Kihara's instigation, enrolled in botany.

Although he was not directly involved in the war, Kimura's life was not easy. There was never enough good food and he disliked the troublesome military drill. Fortunately, the war ended during Kimura's first university year, and he was then able to move to Kihara's laboratory. Civilian life after the war, if anything, was worse than before. On Sundays Kimura regularly visited a cousin who lived in the Kyoto suburbs and had more and better food. He was a quantum physicist, and the scientific talk was a welcome dessert for a much-appreciated meal.

Kimura's first paper in Kihara's lab was an analysis of the distributionof chromosome numbers in successive generations of backcrossing [Kimura, 1950]. Kihara was an enlightened leader and the world of population genetics can be thankful. He recognized Kimura's unusual mathematical ability, so he assigned no specific duties and permitted him to study on his own. Kimura took advantage of this opportunity and began a systematic review of mathematical genetics. Sometimes this necessitated a trip to the library of the University of Tokyo, a long, slow train ride, very hot in the summer. Kimura especially admired Sewall Wright and, since there was no copy machine, he laboriously hand-copied Wright's classic 63-page article [Wright, 1931]. It is fascinating to examine this copy, for Kimura frequently added clarifying notes and already was thinking of alterative derivations and mathematical improvements.

In 1949 Kimura joined the staff of the National Institute of Genetics in Mishima. He immediately plunged into research and the first annual report, issued in 1950, had five of his contributions. Nevertheless, it was a lonely time. No one in Mishima understood the work and few cared for it. Furthermore, for Kimura, Mishima was a rather uninteresting place; he missed the rich cultural life in Kyoto.

One person who did appreciate his work was Taku Komai, who had studied

in the Morgan Drosophila laboratory in the United States. Komai encouraged Kimura to study abroad and with the help of two American geneticists at the Atomic Bomb Casualty Commission Kimura was able to put together enough funds to come to the United States. He had hoped to study with Sewall Wright, but Wright was retiring and not accepting students, so he recommended Iowa State College to study with America's leading animal breeder, Jay L. Lush.

The main thrust of work at Iowa State was analysis of genetic variance, particularly subdivision of epistatic components. Although Kimura understood this work, his heart was not in it. He wanted to work on stochastic processes in the Fisher-Wright tradition. He then asked me if, at the end of his year at Iowa State, he could study at Wisconsin. I was delighted to have a student of such talent and accomplishment. Another plus was that Sewall Wright was moving to Wisconsin. So Kimura moved to Wisconsin and at last had a chance to get well acquainted with his hero and to work on his chosen subject. Curiously, although they often spoke, Wright and Kimura did not work together; their mathematical methods were too different.

The circumstances of my meeting Kimura were quite remarkable. During the sessions of the Genetics Society of America, which that year was held at the University of Wisconsin, I encountered a Japanese person who was had clearly lost his way. Earlier, my student Newton Morton, who was working for the Atomic Bomb Casualty Commission in Japan had discovered Kimura's work and sent me some reprints. I must have been one of a very small handful of people who had ever heard of Kimura, and this immediately started a friendship. He had brought with him a paper, written on the long trip across the Pacific. It was a study of fluctuating selection coefficients, and he had discovered a transformation that converted a very difficult partial differential equation into the well-understood equation of heat diffusion. I helped him with the English and with getting it into print [Kimura, 1954].

The two years that Kimura spent as a graduate student at the University of Wisconsin were astonishingly productive. He wrote several path-breaking papers, while at the same time taking courses, passing language exams, and filling various requirements for the Ph.D., which he received in 1956.

Returning to Japan, he remained at the National Institute of Genetics for the rest of his life, except for a several short visiting appointments throughout Europe and the United States. In particular, he returned several times to the University of Wisconsin, where he and I continued our cooperative research. Likewise, I found several opportunities to work in Japan, so we continued our collaboration for the rest of his life.

Finally, he contracted amyotrophic lateral sclerosis (Lou Gehrig's disease). His last months saw only steady physical deterioration. He died on his 70th birthday anniversary, November 13, 1994. The death was accidental, following a fall, but I think it was merciful, since he had nothing to look forward to but further deterioration.

Kimura has been honored throughout the world, and in particular in his birth-

place, Okazaki. There is a museum and a public statue. For a more detailed biography and account of Kimura's work, see [Crow, 1997].

2 POPULATION GENETICS THEORY

Soon after beginning graduate work at the University of Wisconsin, Kimura wrote one of his most influential papers. He was particularly adept at using diffusion equations. Starting with a population with a specified frequency of two neutral alleles, he worked out the probability of every allele frequency for all future times until one allele is eventually lost from the population. He not only constructed the entire process for the first time, but in doing so he demonstrated the great power and convenience of a diffusion approximation. He soon extended this to three alleles, and on to any arbitrary number. He then added mutation, selection, and migration. His general procedure was to use the Kolmogorov forward equation greatly extending the work of Wright [Wright, 1945]. This work was all done in his first year of graduate work. He was invited to the Cold Spring Harbor Symposium of 1955, where this material and much more was presented [CSH, 1955]. He also introduced an elementary derivation of the Fokker-Planck equation, widely used by physicists but coming into use by evolutionists. The paper was a *tour de force*; but it was doubly difficult to understand, both because of its mathematical complexity and because of Kimura's then-limited skill in English pronunciation. In the discussion Sewall Wright, with his usual generosity, rose to say that only those who had tried such problems could appreciate the enormity of Kimura's accomplishment.

After graduation in 1956, Kimura returned to Japan where he continued studying stochastic processes. In particular, he made use of both the forward and backward forms of the famous Kolmogorov equations [Kimura, 1964]. This enabled him to answer such questions as: the probability of ultimate fixation of a mutant allele with arbitrary initial frequency and selective advantage, the distribution of number of generations required for fixation or loss of a mutant allele, and the number of individuals affected during the time the mutant allele is in transit. Together with his colleague Takeo Maruyama, he discovered a way to obtain a variety of functions of allele frequencies undergoing stochastic evolution [Kimura and Maruyama, 1971].

Fisher's "Fundamental Theorem of Natural Selection" [Fisher, 1930] has long been a subject of controversy. It says that the instantaneous rate of change of fitness is equal to the genic variance of fitness at the time, genic variance being defined by least squares. It is exact only under highly simplified conditions, but as an approximation tells a great deal about how selection works [Crow, 2002]. One of Kimura's accomplishments was to write explicit formulae, taking into account dominance, epistasis, and variable selection coefficients [Kimura, 1958].

One of Kimura's most interesting discoveries was that when there is directional selection in a population, after a few generations it attains a state that he called "quasi-linkage equilibrium". In this state, the amount of linkage disequilibrium

generated almost exactly cancels the epistatic variance. Thus, contrary to the wisdom of the time, a long-time selection program can be predicted more accurately by ignoring epistasis. It vindicated his earlier lack of interest in epistatic components while he was at Iowa State University. For a discussion, see Crow and Kimura [Crow and Kimura, 1970], pages 195-225.

This is only a sample of Kimura's enormous number of accomplishments in mathematical population genetics. The best way to key into Kimura's work is a volume of his major papers that have been reprinted with explanatory introductions by N. Takahata, published in 1994 by the University of Chicago Press [Kimura, 1994].

3 THE NEUTRAL THEORY

In 1968, Kimura presented his block-buster [Kimura, 1968]. This happened in the beginning days of molecular evolution and the pattern of amino acid changes in various evolving proteins was becoming known. The conventional wisdom was that natural selection was the main driving force. Kimura dared to say that most molecular evolution — evolutionary changes at the DNA level — were selectively neutral, driven by the mutation rate and random allele frequency drift. A consequence is that, viewed over a long time period, the rate of neutral evolutionary change is determined by the mutation rate alone.

Kimura's first arguments were not very convincing, but at about the same time King and Jukes [King and Jukes, 1969] reached the same conclusion. There was a great deal of debate, some quite vociferous. To some traditionalists, the idea of neutral evolution was simply unthinkable. Nevertheless, additional evidence in favor of neutrality accumulated.

It should be emphasized that Kimura did not deny that evolution of form and function is driven by natural selection. What he asserted is that this was a small part of change at the DNA level, most of which was evolving neutrally. With time it became clear that in mammals only a small fraction of the DNA codes for proteins; the great bulk is "junk" with no known function. It is reasonable that much of the evolutionary change in this junk is neutral while changes in amino acid composition are more often selected. The jury is still out on the question of just how much DNA change is neutral. Although the neutral theory remains controversial, one point is not in doubt. Being concrete and mathematically simple it leads to testable predictions, and this has been one of its greatest merits. It has had great heuristic value and a neutral prediction is often used as a null hypotheses for testing various models of selection.

It is common in the history of science for mathematics that was developed for its own sake or for a specific subject turns out to have a totally unexpected use. Examples abound, but usually the application comes long after the theory was first developed. Kimura had the good fortune that much of the mathematical theory that he developed in the early days of his career was almost made to order for the neutral theory. Such problems as the probability of fixation of a neutral mutation,

the time till fixation, and the number of individuals in the path to fixation all were immediately applicable to the neutral theory. This made the theory quantitative from the beginning, and permitted many quantitative predictions and tests. It also provided a rational basis for a molecular clock. Remarkably all this happened in one lifetime, Kimura's.

For the rest of his life Kimura devoted most of his energies to finding new approaches to neutral evolution and defending his theory. Each new molecular discovery was examined from this point of view, and often was used as an additional argument. Along the way, he discovered a variety of new mathematical techniques, useful in evolutionary studies and an important part of the permanent repertoire of mathematical population genetics. Kimura summarized his theory, and much more, in a readable book [Kimura, 1983].

4 KIMURA THE MAN

Kimura was widely honored during his lifetime. He received honorary doctorates from the Universities of Chicago and Wisconsin. In addition, he received 14 other high awards; some of the most prestigious are foreign membership in the Royal Society of London and the United States National Academy of Sciences, and the Darwin Medal of the Royal Society.

He wrote well over 150 major papers and several books. His "The Neutral Theory of Molecular Evolution" [Kimura, 1983] was favorably reviewed and widely read. It has become a classic.

While a student, Kimura enjoyed discussing all kinds of subjects: genetics, of course, but also Greek tragedies, music, and science fiction; he especially liked Arthur Clarke. He also enjoyed philosophy, especially Bertrand Russell. He and I had many happy times together, starting with his student days and continuing through his life. He was a delightful companion.

With scientific colleagues and especially advocates of theories contrary to his, Kimura was not so gracious. He could be argumentative, petty, and vindictive. A scientific difference often became personal. In the last decades of his life, defending the neutral theory became an obsession. It was hard to get him to talk about anything else. This singleness of purpose undoubtedly spurred him to more research, but it marred his relations with colleagues and other scientists that he regarded as competitors or opponents. Ironically, the theory stood on its own very well.

In an earlier period, the best-known geneticist in Japan was Hitoshi Kihara. After his retirement, Kimura took his position. Now Kimura has taken a place alongside the great population genetics trinity: R. A. Fisher, J. B. S. Haldane, and Sewall Wright. For Wright's final evaluation of the work of all four, see [Wright, 1988].

BIBLIOGRAPHY

[Crow, 1997] J. F. Crow. Motoo Kimura. *Biographical Memoirs of Fellows of the Royal Society*, 47: 253–265, 1997.

[Crow, 2002] J. F. Crow. Perspective: Here's to Fisher, additive genetic variance, and the Fundamental Theorem of Natural Selection. *Evolution*, 86: 1313-1316, 2002.

[Crow and Kimura, 1970] J. F. Crow and M. Kimura. *An Introduction to Population Genetics Theory. Harper and Row*. New York, 1970.

[CSH, 1955] *Cold Spring Harbor Symposia on Quantitative Biology*. Biology Lab. Cold Spring Harbor, New York. 20: 33–51, 1955.

[Feller, 1950] W. Feller. *An Introduction to Probability Theory and its Applications*. Vol 1, Second edition. John Wiley, New York, 1950.

[Fisher, 1930] R. A. Fisher, *The Genetical Theory of Natural Selection*. Oxford Univ. Press Oxford, 1930. (Complete Variorum Edition, 1999.)

[Kimura, 1950] M. Kimura. The theory of chromosome substitution between two different species. *Cytologia*, 15: 281–284, 1950.

[Kimura, 1954] M. Kimura. Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensity. *Genetics*, 39: 280–295, 1954.

[Kimura, 1958] M. Kimura. On the change of population fitness by natural selection. *Heredity*, 12: 145–157, 1958.

[Kimura, 1964] M. Kimura. Diffusion models in population genetics. *J. Appl. Prob.*, 1: 177–232, 1964.

[Kimura, 1968] M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217: 624–626, 1968.

[Kimura, 1983] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge Univ. Press, Cambridge, 1983.

[Kimura, 1994] M. Kimura. *Population Genetics, Molecular Evolution, and the Neutral Theory. Selected Papers*. University of Chicago Press, Chicago, 1994.

[Kimura and Maruyama, 1971] M. Kimura and T. Maruyama. Some methods for treating continuous stochastic processes in population genetics. *Jap. J. Genetics*, 46: 407–410, 1971.

[King and Jukes, 1969] J. L. King and T. H. Jukes. Non-Darwinian evolution: Random fixation of selectively neutral mutations. *Science*, 164: 788–798, 1969.

[Wright, 1931] S. Wright. Evolution in Mendelian populations. *Genetics* 16: 97–159, 1931.

[Wright, 1945] S. Wright. The differential equation of the distribution of gene frequencies. In *Proc. Natl. Acad. Sci. USA*, 31: 383–389, 1945.

[Wright, 1988] S. Wright. Surfaces of selective value revisited. *Amer. Natur.*, 131: 115–123, 1988.

This page intentionally left blank

Part II
Evolution

This page intentionally left blank

NATURAL SELECTION

Christopher Stephens

Charles Darwin introduced the term "natural selection" in chapter IV of his book *On the Origin of Species* [1859]

> [C]an we doubt (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable variations and the rejection of injurious variations, I call Natural Selection. (p. 80–81)

Before Darwin, the term "selection" was used by plant and animal breeders to indicate a conscious choice to breed organisms with desired characteristics. Darwin used artificial selection as an analogy to understand natural selection, where nature does the picking and choosing.

In contemporary evolutionary theory, "natural selection" is usually defined as occurring if and only if there is heritable variation in fitness [Lewontin, 1970]. For instance, if giraffes with taller necks are more likely to survive than those with shorter necks because they can reach more food sources, then the giraffes with taller necks are *fitter*. If this fitness variation is transmitted to the calves so that giraffes with taller necks are more likely to have taller offspring and shorter giraffes more likely to have shorter offspring, then the trait is said to be *heritable*. If heritable variation in fitness exists, then we should expect, *ceteris absentibus* (all other forces being absent [Joseph, 1980]), fitter traits to increase in frequency and less fit traits to decline. The reason for the *ceteris absentibus* condition is that processes other than natural selection such as random genetic drift, recombination, mutation or migration can also affect evolutionary change. Consequently, heritable variation in fitness (natural selection) is neither necessary nor sufficient for evolution [Endler, 1986].

Following Darwin, biologists often use the term "sexual selection" to refer to selection that is a result of differential mating success, and reserve "natural selection" to refer to other components of fitness, such as viability. This is because some traits enhance fertility but impair viability. The peacock's tail, for instance, is favored by sexual selection because females prefer males with large colorful tails. At the same time, a large colorful tail makes the peacock more likely to be a meal for predators. When natural selection is defined as "heritable variation of fitness" (as above), sexual selection is just a special kind of natural selection.

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

Trying to understand the precise scope and role of natural selection in evolutionary theory generates many questions of philosophical interest. For instance, consider a trait that has evolved by natural selection such as giraffe neck length. Did natural selection favor longer necks because it benefited the *organisms* that had longer necks, because it benefited the *groups* of giraffes that had long necks, because it benefited the *genes* that are (in part) responsible for long necks, or because of some combination of these factors? This is known as the units of selection problem, and will be discussed in a separate entry (see Wilson's chapter on levels of selection in this volume).

We can divide the philosophical issues raised by natural selection into two rough categories. First, there are issues — such as the units of selection problem — that involve understanding the process natural selection itself. Besides the units of selection problem, philosophers have been particularly interested in issues about how to understand and interpret fitness, as well as the question of whether natural selection explains why individuals (as opposed to populations) have the properties that they do. A second set of issues includes questions about the *relationship* between natural selection and other evolutionary processes. Included here are the following sorts of questions: How powerful a force is natural selection compared to the other possible causes of evolutionary change? Are the processes of natural selection, drift, mutation and so on properly understood as forces or causes, or are these factors merely statistical phenomena? I will consider each of these issues in turn.

Fitness

Herbert Spencer [1864] was the first to use the phrase "survival of the fittest" to describe natural selection. Alfred Wallace, who independently discovered and developed the theory of evolution by natural selection [Wallace, 1859], encouraged Darwin to adopt Spencer's phrase because he thought that the term "natural selection" would mislead people into thinking that the process required conscious thought and direction. Darwin did eventually use the phrase "survival of the fittest" [Darwin, 1868] but refused to give up the term "natural selection", and in his later works used the two interchangeably.

Darwin and other biologists' subsequent use of the phrase "survival of the fittest" gave rise to a criticism of the theory of evolution known as the tautology problem. If it is a matter of *definition* that the fittest organisms survive, then in what sense is it explanatory to say that the reason why one group of organisms survived and another did not is because the first group was fitter? As a result of this concern, some philosophers, most famously Popper [1963, later retracted], biologists [Peters, 1976] and various creationists were led to think that this lack of testability was a serious problem with evolutionary theory. Notice, however, that even if some *component* of evolutionary theory is a tautology, this does not mean that evolutionary theory *as a whole* is "untestable" or "true by definition". There is much more to evolutionary theory than the definition of fitness: the mere

definition of fitness does not imply that there actually are heritable variations of fitness, or the nature of these differences, or how they interact with other forces of evolution, and so on. Even if the principle of natural selection is true by definition, just as it might be true by definition that "bachelors are unmarried", this doesn't mean that there can't be empirical ways of finding out about fitness, just as there are empirical ways of finding out who is a bachelor [Sober, 1984].

At any rate, one of the most important developments in responding to the tautology problem was the development and defense of the so-called *propensity interpretation* of fitness [Mills and Beatty, 1979; Brandon, 1990; Sober, 2000]. According to the propensity interpretation, we can think of fitness as a probabilistic dispositional property analogous to the way we think about dispositions such as solubility. Just as a cube of sugar might have a certain probabilistic disposition to dissolve when immersed in a certain liquid, we can think of the fitness of an organism as a probabilistic tendency to survive to adulthood (viability fitness) or a disposition defined by a probability distribution with respect to how many offspring the organism could have (fertility fitness). Suppose that one organism has a .8 chance of having 3 offspring (and a .2 chance of having none) whereas another organism merely has a .7 chance of having 3 offspring (and a .3 chance of having none). This means that each organism has a fertility fitness that is defined by a certain probabilistic propensity known as the *expected* number of offspring. The first organism is fitter than the second, in that it is more likely to have more offspring, but because fitness is a probabilistic propensity, there is no guarantee that the first will have more offspring than the second.

The propensity interpretation identifies fitness with an organism's *expected*, rather than its actual, degree of reproductive success. If fitness were identified with the actual reproductive success of an organism or trait type, it is hard to see how fitness could *explain* the reproductive success of the organism or trait since nothing can explain itself. Defenders of the propensity interpretation argue that the identification of fitness with actual reproductive success commits an operationalist fallacy — defining a term by the way in which it is measured. By thinking of fitness as an expectation, one could argue that fitness explains actual reproductive success the same way that a coin's expectation of heads and tails explains the results of a series of tosses. This suggestion avoids the tautology problem; however, it isn't clear that such a dispositional account is particularly explanatory. If we explain why a group of organisms with one trait survived while another group with a different trait died by saying that the former trait was fitter than the latter, this amounts to saying that what occurred did so because it was more probable than the alternative, which doesn't seem especially illuminating [Sober, 1984].

Other objections have been raised to the propensity interpretation [Rosenberg, 1985; Beatty and Finsen, 1989; Sober, 2001]. One problem is that there is not just a single propensity interpretation of fitness; rather, there are several different ways of understanding fitness as a propensity. For instance, there can be differences between the short and long term fitness of a trait. A given trait might have a high

short-term fitness but a low long-term fitness or vice versa. Another complication is that when a certain type of organism has an offspring contribution that varies solely *between* generations, then the geometrical mean is the appropriate measure of fitness, but when there is variation solely *within* generations, the arithmetical mean is the better measure of evolutionary expectation [Beatty and Finsen, 1989]. It can be difficult to determine whether the variation is distributed one way or the other, and it is also possible that a type will change reproductive strategies over time. There are still many unresolved issues about how to interpret fitness.

What Does Natural Selection Explain?

Sober [1984; 1995] argues that natural selection can explain the frequencies of traits — such as being long-necked — in a population, but doesn't explain why *individual* giraffes have long necks. Selection has a merely negative role of eliminating variants and does not have a "creative" role of explaining why a particular individual has the traits it does. Sober illustrates this with an analogy involving a selection process for admission into a classroom. In order to be admitted to the next grade level, a student must pass an admissions test, which involves being able to read. Suppose children A, B and C pass the admissions test but the other children do not. The selection process (the admissions test) explains why individuals in the next grade level have trait P (being able to read), and why there are no individuals with trait not-P in the next grade level. However, the selection process does not explain why child A has P rather than not-P nor why child B has P rather than not-P, and so on. The selection process determines why the room contains only readers (those with trait P) but does not explain why A is able to read. A's ability to read is presumably explained by appeal to the history of her particular experiences and mental capacities. Analogously, Sober argues that natural selection explains why a population has the *types* of organisms that it does, but it doesn't explain why the *particular* organisms have the traits that they do. Presumably, something like a *developmental* explanation would explain why a particular organism has the traits that it does.

Some philosophers have argued, *pace* Sober, that selection can, in the right circumstances, explain why individuals have the traits that they do [Neander, 1988; 1995a; 1995b; Matthen, 1999; 2002; 2003], while others have defended the view that selection does not explain the traits of individuals [Sober, 1984; 1995; Walsh, 1998; Lewens, 2001; Pust, 2001]. Neander, for example, argues that natural selection can explain the traits of individuals (e.g., why each giraffe has a long neck) by being part of a *cumulative* selection process. In the first stage, selection explains why a certain ancestral population has the traits that it does — e.g., why all of the individuals have genes that produce long necks. Those with genes that coded for short necks died out before reproducing. In the second stage, the traits of the descendents of these giraffes are explained by noting that they inherited their traits from the members of this ancestral population, which was full of long necked survivors. Hence, selection helps explain the traits of these individual

giraffe descendents.

Matthen [1999; 2002; 2003] offers an independent argument in favor of the claim that natural selection can explain the traits of individual organisms. In his scenario, there is also a two-stage process but not one that involves cumulative selection. A single selection event can influence the composition of future generations by increasing the chances that a particular individual will have offspring with a particular trait because a selection event can change the frequencies of genes available. In Matthen's [1999] example, a selection process favors a certain trait C in a population of sheep. As a result, this increases the proportion of sheep with trait C in the population, and makes it more likely that any given sheep will mate with a sheep with trait C. In particular, if we consider a particular sheep P (parent), the *prior* selection event in favor of trait C increases the probability that P will mate with a sheep that has trait C. This means that there could be a particular sheep — let's call it Dolly — that is the offspring of P, whose possession of trait C is in part explained by the prior selection event.

Matthen argues that in order to defend his view, Sober would have to rely on a contentious ontological assumption about origin essentialism — in particular, he must assume that a particular sheep such as Dolly must come from the parents it did come from. Sober needs to distinguish between an individual *like* Polly coming into existence with the traits that Polly has and Polly *herself* coming into existence with the traits that she has. If we help ourselves to this distinction, then someone such as Sober could argue that the prior selection event in Matthen's sheep example only increases the probability that a sheep *like* Polly will have trait C; it does not, however, increase the probably that *Polly* has trait C. This is because in order for it to be Polly the sheep, she must have been born to the particular parents that she had [Kripke, 1972]. Once you have fixed on Polly's particular parents, you have screened off the influence of the prior selection process. Matthen (correctly) doubts that this assumption about origin essentialism is crucial to population genetics, and so concludes that natural selection can explain (in part) why individuals have the traits that they do. Pust [2001] in contrast, argues in favor of origin essentialism.

Forber [2005] argues that there are two related questions that need to be distinguished. First, what is the role of selection in explaining the origin of biological traits and second, what are the implications of this for explaining why particular organisms have the traits that they do? He argues that selection can explain why a trait originates if the trait is a result of a combination of factors. Forber has us imagine two scenarios where there is a haploid organism with two relevant loci. At one locus are alternative alleles A and a and at the other locus are alternatives B or b. In one case, we suppose that the AB combination is the fittest and that the aB and Ab combinations are less fit, and the ab combination the least fit. If the population starts out at 100% ab, we can then appeal to selection to explain why the AB combination arises in the first place. If mutations from a to A and from b to B are relatively rare, then selection can help explain how the population gets to AB. This is because there is a selective advantage to either aB or Ab over ab. so selection increases the probability that an individual (and eventually the entire

population) will get to AB. In the second case, we suppose that AB is again the fittest combination but that aB, Ab and ab are all equally less fit than AB. In this case selection does *not* play a role in explaining why AB first originates in the population. In this latter case, it is only the unlikely combination of mutation and drift that explains the origination of the AB combination. Consequently, selection can explain the origin of biological traits if the right sort of processes occurs. It is not enough that the process is cumulative and it doesn't matter whether sexual reproduction is involved; what is crucial is that a trait is affected by multiple factors.

Forber argues that to resolve the second issue — whether selection can explain why individuals have the traits they do — one must settle certain philosophical issues about explanation that are independent of the purely scientific features of evolutionary biology. As Matthen's sheep example indicates, if selection helps explain why a trait first originates, then it helps explain why *some individuals or other* have that trait. There is, however, a separate debate over whether this explanation about origin also carries over to explain why *particular* individual organisms have such traits.

Defenders of the Neander-Matthen position accept the view that part of what explains why an individual has a particular trait depends on explaining how the trait originated in an ancestral individual. Proponents of the Sober position, on the other hand, argue that a developmental explanation of an individual is sufficient to explain why individuals have the traits that they do — there is no role for what Sober calls a variational explanation. A final resolution of this debate thus depends on sorting out which view of explanation is the most plausible.

Adaptationism

A trait such as giraffe neck length is an *adaptation* for eating leaves high in trees if it is a result of natural selection acting for that end. If, in the relevant ancestral population, there was selection for having longer necks because giraffes with longer necks were more fit than those with shorter necks by virtue of being able to eat more leaves, then we can say that the giraffe's long-neck is an adaptation for eating.

A few remarks about this definition of adaptation. Notice that an adaptation is an *historical* property rather than a comment about a trait's current utility. Maybe long necks are *adaptive* now because they help giraffes see over cars or buildings, but these are not reasons why the trait evolved. A trait can have properties that are not part of why it evolved, such as the fact that a neck has a certain mass. These traits are sometimes known as "spin-off" or "side effects" and there is merely selection *of*, rather than *for*, these properties [Sober, 1984]. A trait can also be favored by natural selection for more than one reason, either simultaneously or in sequence. Giraffe bulls engage in "necking displays" to exhibit dominance, and so long necks might also be an adaptation for this behavior. In some cases, a trait such as having a long neck might evolve for one reason (to reach high leaves)

and then be favored by natural selection for another reason (dominance displays). Gould and Vrba [1982] refer to this phenomenon as an "exaptation" and argue that traits such as a long neck are really only adaptations for why they were *first* favored by natural selection, and are merely exaptations for the subsequent behavior. I see no reason, however, to restrict the term adaptation in such a way. Surely a trait can be an adaptation both for why it was first selected *and* for why it is maintained by selection [Sterelny and Griffiths, 1999]. At the same time, Gould and Vrba are right to remind us that one cannot simply assume that the reason that a trait is favored *now* or in the recent past is the same reason that it was favored by selection in the more distant past.

The central debate about adaptationism concerns the *power* and *scope* of natural selection. Everyone agrees that traits like eyes and wings were favored by natural selection for some reason, though there are disputes about exactly why such traits were favored. Biologists also agree that natural selection is an important process in evolution. There is considerable disagreement, however, about the significance of non-selective factors such as random genetic drift and the role of mechanical, developmental, genetic, and other sorts of constraints. Is natural selection powerful enough so that it can overcome constraints and other non-selective processes? Can model builders effectively ignore factors such as drift, at least in most circumstances?

Although debates about the importance of natural selection go back to Darwin and Wallace, the contemporary discussion is in large part a reaction to Gould and Lewontin's [1979] famous paper "The spandrels of San Marco and the Panglossian paradigm". In their paper, Gould and Lewontin raise two main kinds of objections to adaptationism — empirical and methodological. On the empirical side, they claim that many biologists tend to confuse adaptiveness (current utility) with adaptation by assuming that a trait must have been selected in the past for what it does well now. They also charge that adaptationists tend to "atomize" organisms — thinking that each trait of an organism can be assumed to evolve independently from its other traits. The number of digits on one hand should not be understood as a trait that evolves independently from the number of digits on our other hand. This is an extreme case, of course, but anti-adaptationists tend to think that there are many such constraints and correlations between traits, whereas adaptationists tend to think there are not.

In addition to disagreement about empirical issues, adaptationists and their critics have methodological disputes as well. Gould and Lewontin charge that adaptationists, when faced with the failure of one adaptive explanation, tend to immediately come up with an alternative adaptationist explanation rather than look to a non-adaptationist explanation. In particular, Gould and Lewontin want to encourage biologists to pursue the idea of a body plan (bauplan) and to take seriously physiological constraints that limit the power and scope of natural selection.

Gould and Lewontin's most important methodological criticism of adaptationism is that adaptationists tend to make up "just-so stories" — how-possibly ex-

planations that aren't supported by much evidence. It is easy to make up a story about how a trait *might* have been favored by natural selection and much harder to get evidence in favor of a particular adaptationist explanation. It may seem obvious that long necks in giraffes are an adaptation for reaching leaves high in trees, but if it turns out that long necks evolved in an environment without tall trees, this explanation would be disconfirmed.

Some adaptationists, such as Richard Dawkins [1982; 1983] and Daniel Dennett [1995], defend their view by pointing out that natural selection is the only process capable even in principle of explaining adaptive complexity. Dawkins argues, for instance, that Lamarckian methods of use and disuse plus the inheritance of acquired characteristics cannot explain even in principle how a trait such as the eye could have evolved, since there is no way that the use of a proto-eye could change its properties to make it more transparent and hence better adapted. It should be noted that however indispensable natural selection might be in order to explain adaptive complexity, this response does not answer the methodological challenges raised by Gould and Lewontin.

There is considerable dispute about how exactly to formulate adaptationism [Maynard Smith, 1978; Lewontin, 1979; Dawkins, 1982; Orzack and Sober, 1994; Dennett, 1995; Brandon and Rausher, 1996; Sober, 1996; Amundson, 2001; Godfrey-Smith, 2001]. The adaptationism debate has often involved caricatures and misrepresentations of the opposing side. Lewontin (1979), for instance, describes adaptationism as "that approach to evolutionary studies which assumes without further proof that all aspects of the morphology, physiology and behavior of organisms are adaptive optimal solutions to problems." As described, this is probably not a position held by anyone, and makes adaptationism trivially false. There is, however, broad agreement that there are at least two different kinds of adaptationism. On the one hand there is what is sometimes called Empirical Adaptationism [Sterelny and Griffiths, 1999; Sober, 2000; Godfrey-Smith, 2001]. *Empirical Adaptationism* is usually defined as the view that natural selection is the most or only important cause of most traits in most populations. The thesis may be stated more or less strongly, depending on how powerful and how pervasive one claims that natural selection is. Anti-adaptationists, by contrast, often emphasize the role of constraints in limiting the power of natural selection.

Consider a case of heterozygote superiority. Suppose we have a diploid organism that can have either a A or a allele at a given locus. That is, each organism in the population has either a aa or AA homozygote or a Aa heterozygote. Suppose further that the heterozygote codes for the fittest trait. Sickle-cell anemia is a famous case of this phenomenon. Humans with one copy of the a allele acquire a resistance to malaria but still have healthy cells. People with two copies of the a allele develop a serious disease known as sickle-cell anemia. AA homozygotes have neither the resistance to malaria nor the anemia. The heterozygote is the fittest trait, but heterozygotes don't "breed true", since half of the time when two heterozygotes mate their offspring will *not* be Aa. This means that natural selection cannot cause the fittest trait to sweep to 100% of the population because

of this genetic constraint.

Adaptationists tend to think that cases of heterozygote superiority are rare — there is a good reason that most biology textbooks use the sickle-cell example, since such cases, they argue, are unusual. Anti-adaptationists, in contrast, worry that genetic constraints might be more common. The fittest trait can also fail to evolve where heterozygotes are *inferior* to both homozygotes. In such cases, the population might evolve to 100% of the homozygote that is not the most fit. Unlike the case of heterozygote superiority, there would not be a phenotypic polymorphism as with the sickle-cell anemia case. Once the population is at equilibrium, the evolutionary process may have destroyed the information about whether the other homozygote was the fitter alternative.

Another kind of constraint is developmental — fruit flies have only a limited number of eggs and can have either high fecundity early in life or late in life, but not both. It might in some sense be optimal to have a high fecundity both early and late in life, but it appears that this not an option because of constraints. Many traits — e.g., nipples in males — are developmental byproducts, rather than adaptations.

Adaptationists also worry about the extent of pleiotropy, in which one gene leads to two different phenotypic effects. If one of these effects is harmful, the case is known as antagonistic pleiotropy. Dawkins tends to think that the power of natural selection will overcome such constraints. He writes "if a mutation has one beneficial effect and one harmful one, there is no reason why selection should not favour modifier genes that detach the two phenotypic effects, or that reduce the harmful effect while enhancing the beneficial one" [Dawkins, 1982, 35].

The second major form of adaptationism is methodological adaptationism. *Methodological adaptationism* is the view that adaptationism is an indispensable research tool for finding out about whether a trait is an adaptation. Sober states the position a bit more strongly "The only way to find out whether an organism is imperfectly adapted is to describe what it would be like if it were perfectly adapted." [Sober, 1996, 54]. Stephens and Krebs [1986] say something similar — the only way that one can get information about constraints is via an optimality model. Amundson [2001] argues, however, that one can sometimes get information about constraints from facts about developmental biology alone and not by an optimality model. Still, as I discuss in more detail below, adaptationists can claim that theirs is a position only about the power of selection acting on phenotypes that are actually generated. Methodological adaptationism, though somewhat contentious, is less controversial than empirical adaptationism.

Critics such as Gould and Lewontin [1979] argue that adaptationist explanations are too easy to make up. Biologists have responded in part by making theories with predictions that are harder to confirm. Biologists have two major tools for confronting the methodological challenges associated with adaptationism [Kitcher, 1985; Reeve and Sherman, 1993; Sterelny and Griffiths, 1999; Sober, 2000]. On the one hand, as Maynard Smith [1978] has emphasized, biologists can and should develop optimality models that make precise, quantitative predictions.

If the model makes precise predictions, it is no longer a trivial matter to make up a theory or model that fits the facts. Furthermore, one can often build into one's model various kinds of constraints, thereby dealing with various concerns about certain sorts of mechanical and engineering constraints. It should also be noted that *non-adaptationist* explanations are easy to make up too, which means that precise models are needed to test *alternatives* to adaptationism as well.

Of course, critics of adaptationism point out that this still leaves room for various *ad hoc* maneuvers that making testing adaptationist claims more difficult. Dawkins discusses such a case in considering an appeal to a "time lag" between present and past conditions to explain away data that does not fit a model. Lack [1966] used this approach when discussing optimal clutch size. In a case where the data didn't match up to his theory, Lack appealed to the claim that the conditions under which the clutch size evolved were different from the present circumstances, and so there had not yet been time for the trait to evolve to meet the new environmental conditions. Dawkins claims that although it is post hoc, it is at least still testable in principle. Information about the food conditions in the time period during which the trait evolved would help resolve the issue.

The second major tool that adaptationists avail themselves to is the comparative method [Harvey and Pagel, 1991; Sterelny and Griffiths, 1999; Sober, 2000]. The main idea behind the comparative method is to examine correlations in many species between two more traits or between a trait and some environmental variable. For instance, if increasing neck length is correlated with environments that have tall trees, this correlation can be studied in close relatives of the giraffe to get more data. In some cases, one discovers historical facts that undermine an otherwise plausible adaptationist hypothesis. Some biologists thought that the low birth weight found in bears was a by-product of the evolution of hibernation. It turns out, however, that low birth weight evolved *before* hibernation, and appears on branches of the evolutionary tree in which hibernation never evolved [McKitrick, 1993; Sterelny and Griffiths, 1999].

Peter Godfrey-Smith [2001] argues that there is a third kind of adaptationism, in addition to its empirical and methodological versions. According to Godfrey-Smith, *explanatory adaptationism* is the claim that the most important question in evolutionary biology is how to explain apparent design, and that natural selection is the "big answer" to this question. This is distinct from empirical adaptationism because one could think that natural selection is the answer to the "big question" about apparent design without thinking that natural selection is particularly pervasive or powerful in overcoming constraints. Godfrey-Smith claims that explanatory adaptationism is the kind of adaptationism responsible for the most difficult conceptual problems surrounding the adaptationism debate. He suggests that Dawkins and Dennett are clearly explanatory adaptationists and may not be empirical adaptationists. Dawkins [1983, 16], for instance, defends explanatory adaptationism when he says that explaining adaptive complexity is the main task of any theory of evolution. He also says that "I would not presume to try to convert any of these people to my point of view" (Dawkins 1983, 16), which suggests

that this may not be an issue which is resolvable by any sort of straightforward epistemic means.

While Godfrey-Smith is right to see biologists such as Dawkins (and perhaps philosophers such as Dennett) as defending explanatory adaptationism, it isn't clear that this thesis is a focus of the adaptationism debate. Amundson [1994] develops a similar set of distinctions, except that he is more focused on the particular criticism that adaptationists ignore the significance of developmental constraints. According to Amundson, adaptationists tend to see developmental considerations as a kind of constraint on adaptation. On this picture developmental constraints are on a par with genetic constraints, random genetic drift, and any other factors that "get in the way" of natural selection. If natural selection is powerful enough, it will overcome these other factors and processes.

However, Amundson is concerned that we see the developmental challenge in a more radical way. Developmental constraints are not processes that constrain natural selection the way that drift, mutation, migration, and so on do since these can all be represented in population genetics models along with selection to determine the relative importance of these processes in the evolution of a particular trait in a population. Drift is a theoretical option in the standard approaches to population genetics; even when its *significance* is denied, there is no problem in understanding how it could make a difference in the models, at least theoretically. Development, on the other hand, has been "black-boxed" and not as well integrated. Amundson claims that many fans of developmental constraints see something misguided about the traditional picture in which drift, selection and development are all on a par. Drift and selection share a common vocabulary and mathematical framework, but development does not.

Godfrey-Smith [2001] points to Dawkins' [1982] claim that the neutralism debate is orthogonal to the kind of adaptationism that he (Dawkins) wants to defend as evidence that Dawkins is not an empirical adaptationist. According to Dawkins, the adaptationism debate is concerned with non-molecular traits. He writes that "given that we are dealing with a phenotypic effect big enough to see and ask questions about, we should assume that it is the product of natural selection." So, according to Dawkins, the neutrality controversy is really a secondary issue.

While I agree that much of the adaptationism debate is orthogonal to the debate about neutralism, I don't think this sort of passage is sufficient evidence that Dawkins is not (or perhaps not) an empirical adaptationist. The right sort of attitude to take is to note that there are really two separate controversies here, one about neutralism that is at the molecular level, and a largely distinct debate about adaptations at the macroscopic level. Dawkins can still be viewed as an empirical adaptationist because of his belief that due to the richness of variation natural selection will tend to overcome pleiotropy and various sorts of genetic constraints.

Like Godfrey-Smith [2001], Amundson [1994] points out that defenders of developmental constraints have a different view about what the primary explanandum of evolutionary theory is. Standard neo-Darwinian biology views adaptations

("apparent design") as the primary explanandum, whereas developmental biologists often view the diversity of organic forms as the primary explanandum. On the developmentalist picture, the diversity of organic forms is best explained by constraints-on-form, rather than constraints-on-adaptation.

Amundson points out that there are two different kinds of dispute here — one for each part of the two-stage process involving variation and selection. The first stage concerns the production of heritable variation, whereas the second concerns the sorting of that variation. Natural selection only operates at the second part of this process. The debate between advocates of selection and advocates of drift occurs in this second part of the process, whereas advocates of developmentalism claim that it is at the first part of this process where constraints play a significant role. In other words, development may or may not constrain what the actual variations are in a population, whereas other sort of constraints (such as the genetic ones mentioned previously) operate on the already existing variation.

Amundson then uses this to draw a distinction between "soft" and "hard" adaptationism in which the soft adaptationists allow for the significance of developmental constraints on the process of determining what the variants will be. According to *soft adaptationism*, developmental constrains-on-form must result in a constraint-on-adaptation. *Hard adaptationism*, on the other hand, claims that developmental constraints can be overcome by the power of natural selection, and so constraints-on-form are not constrains-on-adaptation.

Amundson also notes that some constraints-on-form can actually increase certain adaptive possibilities. For instance, because of certain developmental constraints there will be certain correlated changes in form — it is not necessary that both the left hand and the right hand have to independently be the target of selection for them both to change. This means that adaptationists such as Stephens and Krebs (1986) are wrong in thinking that constraints must be defined as restricting adaptations.

Reeve and Sherman [1993] claim that adaptationism is consistent with the existence and importance of constraints, though Amundson points out that there are different kinds of constraints, some of which conflict with adaptationism. Whether adaptationism is in tension with constraints depends on the kind of constraints – adaptationism is genuinely in conflict with there being a large number of genetic constraints. It is not, however, necessarily in conflict with mechanical constraints, such as those that occur when leg length must be traded off with leg strength [Maynard Smith, 1978].

Although Amundson and Godfrey-Smith are raising interesting issues here, it isn't clear how central they are to the adaptationism debate. It is hard, for example, to make sense of the claim that natural selection is more powerful or "does more" in a case where it chooses among many actual variations compared to a case where, due to developmental constraints, there are only a couple of alternate phenotypes to act on. This is precisely because, as Amundson and others point out, natural selection operates at a different point in the process than the developmental constraints. To resolve this issue would require determining the extent to

which the adaptationism debate is about how nature restricts the space of options for natural selection to act on as opposed to how other sorts of constraints (such as genetic ones) restrict natural selection once it has options to act on.

Finally, it is also worth mentioning that empirical adaptationism is best thought of as a research programme (see [Orzack and Sober, 1994; Sober, 2000]). It is possible to test both particular adaptationist (and non-adaptationist) hypotheses as well as general claims about the long run. Particular adaptationist claims are testable if they are stated with enough precision, and whether or not most traits in most populations are primarily a result of natural selection is something that can be tested by accumulating data about many individual cases.

Forces and Causes

Biologists [Roughgarden, 1979; Gillespie, 1998] and philosophers [Sober, 1984] often talk about natural selection, random genetic drift, mutation and other evolutionary processes as forces and causes. Sober, for example, uses the Hardy-Weinberg equation to formulate one of two versions of a "zero-force' law that states what will happen to gene and genotype frequencies if none of the standard processes such as natural selection, mutation, migration and drift are at work in a population. We can speak of natural selection *causing* the trait of having a long neck to evolve in the giraffe lineage. We can talk about how mutation and drift are forces that are *equal in strength*, so that they *counteract* the effect of one another [Gillespie, 1998], or we can say that in small populations drift is a *more important* force than selection. Population genetics has models in which selection *opposes* mutation and models in which mutation operates in the *same direction* as selection, so it looks like these evolutionary processes can have directions and strengths, just like forces.

Other biologists and philosophers have expressed doubts about understanding these processes as forces [Endler, 1986, Walsh *et al.*, 2002] or as either forces or causes [Matthen and Ariew, 2002]. They raise several sorts of doubts about the Newtonian analogy, and prefer instead to think of natural selection as a purely statistical theory. We will consider a few of their objections here, focusing on the processes of natural selection and random genetic drift. They argue that notions such as fitness appear to be statistical summaries rather than causes, and argue that the traditional view exaggerates the degree to which evolutionary theory is analogous to Newtonian physics. They also argue that thinking of drift as a force or a cause has some peculiar consequences. Let us take each of these issues in turn.

To what extent can evolutionary theory be understood as a theory of forces analogous to those in Newtonian physics? First of all, it is important to recognize that the possible evolutionary processes of selection, drift, mutation and so on occur against various background theories about, e.g., how heredity works [Sober, 1984; 1996; Matthen and Ariew, 2002; Stephens, 2004]. It does not make sense to talk about natural selection operating alone, in the absence of a system of

inheritance. So this is one important difference with a Newtonian zero-force law.

Secondly, several biologists and philosophers have pointed out that fitness is not well understood as a causal processes; rather, it is more like the actuarial notion of overall life expectancy [Fisher, 1930; Sober, 1984; Walsh *et al.*, 2002; Matthen and Ariew, 2002; Stephens, 2004]. Overall life expectancy is not a *cause* of how long one lives, it is instead merely a statistical summary of information about how a number of possible causes might affect one's survival. Similarly, an organism's viability fitness is not a cause of its survival (or death); rather, it is a statistical summary of the possible causes that might affect an organism's viability, weighted by their chances of occurring. The fact that a particular cockroach (or a trait type in a population) has a .8 chance of surviving to adulthood is not what causes the cockroach to survive (if it does). It is a reflection of a weighted summary of possible factors — many of which never come to pass - that might kill the cockroach.

Critics of the traditional view think that natural selection is like the notion of fitness or life expectancy — it is not a force or even a cause; rather, it is merely a statistical summary of various life and death events that occur in a population. Endler [1986, 29], for example, compares natural selection to erosion. He claims that natural selection is the *result* of heritable variation in fitness, rather than identical to it or a cause of it. He compares natural selection to erosion, which is a result of variance in resistance to weathering. Endler also objects that natural selection is disanalogous to force in physics, since it cannot be decomposed into something analogous to a mass and acceleration. Endler allows that natural selection can be a cause — it is, however, merely an effect of heritable variation of fitness, but it is a cause that can lead to genetic change in populations.

Defenders of the traditional view often reply by arguing that there is an important distinction between fitness and natural selection. Fitness may not cause an organism to survive or reproduce, but to say that one organism is fitter than another is not to say why it is fitter. Selection is supposed to be a causal notion that explains *why* one trait is fitter than another. Furthermore, terms such as 'natural selection' and 'drift' can be used in two ways — to talk about a process, or to talk about the result of a process. The traditionalists maintain, while the critics deny, that the process notions of natural selection and drift can be understood causally.

Another major source of doubt comes from thinking about drift as a force. First, it should be noted that it is common ground in this debate that it does not make sense to talk about how much drift and selection contribute in an individual case. This makes no more sense that it does to decompose the effects of the bias of a coin and the number of times it is tossed in determining why a particular flip of the coin came up heads or tails. Drift, if it is to be understood as a force or cause, is, as Sober says "a force of a different color" [Sober, 1984, 117], and can only be understood as a population level force or cause.

Critics of the traditional view point out that it is puzzling to think about drift as a force or cause by thinking about the following kind of example. Imagine that there are two scenarios, each with two organisms, one having trait T_1 which is fitter than the other organism, which has trait T_2. In the first scenario, the organism

with T_1 (the fitter trait) survives and the second organism (with T_2) dies. In the second scenario, lightning kills the fitter organism with T_1 and the organism with T_2 survives. It would seem incoherent to say that selection operates in the first scenario while drift operates in the second.

This is correct, but defenders of the traditional view accept this consequence. One cannot say how much drift and selection each contribute in an individual case. The way to determine if drift has had an effect is to compare populations — either two actual populations with different sizes or one actual population with a hypothetical population of infinite size (i.e., where there is no drift). It is natural to think of a population's effective size as one of the causes of why traits in the population evolve. If its size were much bigger or smaller, we would *expect* different frequencies of genes or genotypes to evolve.

The debate about how to interpret natural selection, drift, and other factors will doubtless continue, especially since the disagreements depend in part on disputes about the nature of causation that lie in the background. We have seen that there are a number of issues about natural selection that have interesting philosophical dimensions, and doubtless there are many others.

BIBLIOGRAPHY

[Amundson, 1994] R. Amundson. Two concepts of constraint: Adaptationism and the challenge from developmental biology. *Philosophy of Science* 61: 556-578, 1994.

[Amundson, 2001] R. Amundson. Adaptation and Development: on the lack of common ground. In *Adaptationism and Optimality*, S. Orzack and E. Sober, eds. pp. 303–334. Cambridge University Press, 2001.

[Beatty and Finsen, 1989] J. Beatty and S. Finsen. Rethinking the Propensity Interpretation: a peek inside Pandora's box. In *What the Philosophy of Biology Is*, M. Ruse, ed., pp. 17–30. Kluwer Academic Publishers, 1989.

[Brandon, 1990] R. Brandon. *Organism and Environment*. Princeton University Press, 1990.

[Brandon and Rausher, 1996] R. Brandon and M. Rausher. Testing adaptationism: A comment on Orzack and Sober. *American Naturalist* 148: 189-201, 1996.

[Darwin, 1859] C. Darwin. *On the Origin of Species*, 1859. *A facsimile of the First Edition*, E. Mayr, ed. Cambridge: Harvard University Press, 1964.

[Darwin, 1868] C. Darwin. *The Variation of Plants and Animals under Domestication, 2 vols.* New York: Orange Judd, 1868.

[Dawkins, 1982] R. Dawkins. *The Extended Phenotype*. Oxford: W. H. Freeman, 1982.

[Dawkins, 1983] R. Dawkins. Universal Darwinism. In *Evolution from Molecules to Man*, D. S. Bendall, ed., pp. 403–425. Cambridge University Press, 1983.

[Dennett, 1995] D. Dennett. *Darwin's Dangerous Idea: evolution and the meanings of life*. Simon and Schuster, 1995.

[Endler, 1986] J. A. Endler. *Natural Selection in the Wild*. Princeton: Princeton University Press, 1986.

[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press, 1930.

[Forber, 2005] P. Forber. On the explanatory roles of natural selection. *Biology and Philosophy* 20: 329-342, 2005.

[Gillespie, 1998] J. Gillespie. *Population Genetics: A Concise Guide*. Baltimore: Johns Hopkins, 1998.

[Godfrey-Smith, 2001] P. Godfrey-Smith. Three kinds of adaptationism. In *Adaptationism and Optimality*, S. Orzack and E. Sober, eds, pp. 335–357. Cambridge University Press, 2001.

[Gould and Lewontin, 1979] S. Gould and R. Lewontin. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B* 205: 581-598, 1979.
[Gould and Vrba, 1982] S. Gould and E. Vrba. Exaptation: A missing term in the science of form. *Paleobiology* 8: 4-15, 1982.
[Harvey and Pagel, 1991] P. H. Harvey and M. Pagel. *The comparative method in evolutionary biology*. Oxford: Oxford University Press, 1991.
[Joseph, 1980] G. Joseph. The many sciences and the one world. *Journal of Philosophy*, 77:773–791, 1980.
[Kitcher, 1985] P. Kitcher. *Vaulting Ambition: sociobiology and the question for human nature*. MIT Press, 1985.
[Kripke, 1972] S. Kripke. *Naming and Necessity*. Cambridge: Harvard University Press, 1972.
[Lack, 1966] D. Lack. *Population Studies of Birds*. Oxford: Oxford University Press, 1966.
[Lewens, 2001] T. Lewens. Sex and selection: A reply to Matthen. *British Journal for the Philosophy of Science* 52: 589-598, 2001.
[Lewontin, 1970] R. Lewontin. The units of selection. *Annual Review of Ecology and Systematics* 1: 1-14, 1970.
[Lewontin, 1979] R. Lewontin. Sociobiology as an adaptationist program. *Behavioral Science* 24: 5-14, 1979.
[Matthen, 1999] M. Matthen. Evolution, Wisconsin style: selection and the explanation of individual traits. *British Journal for the Philosophy of Science* 50: 143-150, 1999.
[Matthen, 2002] M. Matthen. Origins are not essences in evolutionary systematics. *Canadian Journal of Philosophy* 32: 167-182, 2002.
[Matthen, 2003] M. Matthen. Is sex really necessary? and other questions for Lewens. *British Journal for the Philosophy of Science* 54: 297-308, 2003.
[Matthen and Ariew, 2002] M. Matthen and A. Ariew. Two Ways of Thinking about Fitness and Natural Selection. *Journal of Philosophy* 119: 55-83, 2002.
[Maynard Smith, 1978] J. Maynard Smith. Optimization theory in evolution. *Annual Review of Ecology and Systematics* 9: 31-56, 1978.
[McKitrick, 1993] M. McKitrick. Phylogenetic constraint in evolution: has it any explanatory power? *Annual Review of Ecology and Systematics* 24: 307-330, 1993.
[Mills and Beatty, 1979] S. Mills and J. Beatty. The propensity interpretation of fitness. *Philosophy of Science* 46: 263-288, 1979.
[Neander, 1988] K. Neander. What does natural selection explain? A correction to Sober. *Philosophy of Science* 54: 422-426, 1988.
[Neander, 1995a] K. Neander. Pruning the tree of life. *British Journal for the Philosophy of Science* 46: 59-80, 1995.
[Neander, 1995b] K. Neander. Explaining complex adaptations: A reply to Sober's 'reply to Neander'. *British Journal for the Philosophy of Science* 46: 583-587, 1995.
[Orzack and Sober, 1994] S. Orzack and E. Sober. Optimality models and the test of adaptationism. *American Naturalist* 143: 361-380, 1994.
[Peters, 1976] R. Peters. Tautology in evolution and ecology." *American Naturalist* 110: 1-12.
[Popper, 1963] K. Popper. *Conjectures and Refutations*. London: Hutchinson, 1963.
[Pust, 2001] J. Pust. Natural selection explanation and origin essentialism. *Canadian Journal of Philosophy* 31: 201-220, 2001.
[Reeve and Sherman, 1993] H. Reeve and P. Sherman. Adaptation and the goals of evolutionary research. *Quarterly Review of Biology* 68: 1-32, 1993.
[Rosenberg, 1985] A. Rosenberg. *The Structure of Biological Science*. Cambridge University Press, 1985.
[Roughgarden, 1979] J. Roughgarden. *Theory of Population Genetics and Evolutionary Ecology*. New York: Macmillan, 1979.
[Sober, 1984] E. Sober. *The Nature of Selection: evolutionary theory in philosophical focus*. Cambridge: MIT Press, 1984.
[Sober, 1995] E. Sober. Natural selection and distributive explanation: A reply to Neander. *British Journal for the Philosophy of Science* 46: 384-397, 1995.
[Sober, 1996] E. Sober. Evolution and Optimality – Feathers, Bowling Balls, and the Thesis of Adaptationism. *Philosophical Exchange* 26: 41-57, 1996.
[Sober, 2000] E. Sober. *Philosophy of Biology, 2^{nd} edition*. Westview Press, 2000.

[Sober, 2001] E. Sober. The two faces of fitness. In *Thinking About Evolution: Historical, Philosophical and Political Perspectives*, R. Singh, D. Paul, C. Krimbas, and J. Beatty, eds., pp. 309–321. Cambridge University Press, 2001.

[Stephens, 2004] C. Stephens. Selection, Drift, and the "Forces" of Evolution. *Philosophy of Science* 71: 550-570, 2004.

[Stephens and Krebs, 1986] D. W. Stephens and J. R. Krebs. *Foraging Theory*. Princeton: Princeton University Press, 1986.

[Sterelny and Griffiths, 1999] K. Sterelny and P. Griffiths. *Sex and Death: An introduction to Philosophy of Biology*. Chicago and London: The University of Chicago Press, 1999.

[Spencer, 1864] H. Spencer. *The Principles of Biology*. London: William and Norgate, 1864.

[Wallace, 1859] A. R. Wallace. On the tendency of varieties to depart indefinitely from the original type.*Journal of the Proceedings of the Linnean Society (Zoology)* 3, 1859. Reprinted in *The Collected Papers of Charles Darwin*, vol. 11: P. H. Barrett, ed., pp. 10–19, 1977.

[Walsh, 1998] D. M. Walsh. The scope of selection: Sober and Neander on what natural selection explains. *Australasian Journal of Philosophy* 76: 250-264, 1998.

[Walsh *et al.*, 2002] D. M. Walsh, T. Lewens, and A. Ariew. The Trials of Life: natural selection and random drift. *Philosophy of Science* 69: 452-473, 2002.

This page intentionally left blank

NEUTRALISM

Anya Plutynski

In 1968, Motoo Kimura submitted a note to *Nature* entitled "Evolutionary Rate at the Molecular Level", in which he proposed what has since become known as the neutral theory of molecular evolution. This is the view that the majority of evolutionary changes at the molecular level are caused by random drift of selectively neutral or nearly neutral alleles. Kimura was not proposing that random drift explains all evolutionary change. He does not challenge the view that natural selection explains adaptive evolution, or, that the vertebrate eye or the tetrapod limb are products of natural selection. Rather, his objection is to "panselectionism's intrusion into the realm of molecular evolutionary studies". According to Kimura, most changes at the *molecular level* from one generation to the next do not affect the fitness of organisms possessing them. King and Jukes [1969] published an article defending the same view in *Science*, with the radical title, "Non-Darwinian Evolution", at which point, "the fat was in the fire" [Crow, 1985b].

The neutral theory was one of the most controversial theories in biology in the late twentieth century. On the one hand, the reaction of many biologists was extremely skeptical; how could evolution be "non-Darwinian"? Many biologists claimed that a "non-Darwinian" theory of evolution was simply a contradiction in terms. On the other hand, some molecular biologists accepted without question that many changes at the molecular level from one generation to the next were neutral. Indeed, when King and Jukes' paper was first submitted, it was rejected on the grounds that one reviewer claimed it was obviously false, and the other claimed that it was obviously true [Jukes, 1991].

Why were some biologists so skeptical and others so nonchalant about the neutral theory? Why was the neutral theory so controversial? What evidence and argument was originally offered on behalf of the theory? How are tests of the theory carried out, and have any of them been decisive? Finally, what is meant by the claim that "drift" operates at the molecular level, independently of change in frequency of phenotypic traits from one generation to the next? What is "drift" in the context of the neutral theory, and how, if at all, is it distinct from drift operating at higher levels in evolution? What are the implications of neutrality at the molecular level, if any, for debates over the prevalence and explanation of adaptation? This short essay will address the above questions.

1 THE NEUTRAL THEORY: SOME HISTORICAL BACKGROUND

What it means to be "neutralist" has changed over the course of the history of evolutionary biology. An uncontroversial sense of "neutralism" is the claim that

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

many *phenotypic* traits have no effect on an organism's fitness. The organism's *phenotype* consists of its physical and behavioral characters or traits – from height to color of plumage. Clearly, not all phenotypic traits have an impact on an organism's capacity to survive or reproduce. Darwin took note of this fact in the sixth edition of the *Origin*:

> I am inclined to suspect that we see, at least in some [cases], variations which are of no service to the species, and which consequently have not been seized on and rendered definite by natural selection. Variations neither useful nor injurious would not be affected by natural selection, and would be left either a fluctuating element, as perhaps we see in certain polymorphic species, or would ultimately become fixed ... We may easily err in attributing importance to characters, and in believing that they have been developed through natural selection; ... many structures are now of no direct use to their possessors, and may never have been much use to their progenitors. [Darwin, 1872]

While it is uncontroversial that there are some phenotypic traits that have no effects on fitness, there has been considerable controversy in the history of evolutionary biology over what *proportion* of phenotypic traits were subject to selection. In other words, there has been a long controversy over whether, and to what extent, chance and accident has played a role in the distribution of phenotypic traits in populations. In the early twentieth century, there were some who thought that many if not most traits were a product of chance, not selection. Gulick [1888], and later, Wright [1931] argued that many traits differentiating local populations (polymorphisms) may be due to drift, or random fixation of traits independent of their selective value. While they argued that selection surely played a role in the differentiation and adaptation of species, they claimed that many polymorphic traits may have simply been a by-product of isolation and sampling error, or, the random sampling of certain types of individuals versus other types from one generation to the next (See [Provine, 1986] for a discussion). In the 1930's and 40's, many biologists accepted the view that quite a few phenotypic traits were due to drift. In the 1950's and 60's however, there was somewhat of a sea change in favor of selectionist views. In part, this was because of the discovery that some phenotypic traits formerly regarded as neutral were in fact selected for (blood groups). (See [Crow, 1985], for a discussion).

While most evolutionists were more or less pan-selectionists[1] in the 1950's and 60's, there was some controversy over what to expect at the genetic level, if indeed, selection was the major factor in phenotypic evolution. At one extreme, proponents of what Dobzhansky coined the "classical view", held that the relentless action of selection should make the genetic material relatively uniform, or homozygous. Deleterious mutations are regularly eliminated, proponents of this

[1] "Pan-selectionism" is the view that all or almost all traits were at some point shaped by natural selection.

view maintained, and occasionally, favorable mutations arise that are eventually incorporated into the population. At the other extreme, the "balance" theorists suggested that genetic polymorphisms might be held in "balance" by selection (For a discussion of this debate, see [Dietrich, 1994]). In other words, it was not obvious in the 1960s, even if selection was the major factor at work in evolving populations, whether the genetic material should as a result be uniform, or how uniform it should be.

"Neutralism", in Kimura's sense, is distinct from both senses mentioned above; i.e., he was not concerned with whether or to what extent, phenotypic evolution was neutral, or whether, assuming selection was in operation, the corresponding genetic material should be heterozygous or homozygous. Rather, he was concerned with a different "level" at which evolution is going on – the molecular level. According to Kimura, "Neutralists claim that the amino acid and nucleotide changes that accumulate within the species in the course of evolution are mainly due to random fixation of selectively neutral mutants" [Kimura, 1976, 152]. Kimura's paper was brief and elegant. The argument was as follows. First, Kimura reports on the rate of amino-acid substitution in hemoglobin, triosephosphate dehydrogenase, and cytochrome c in mammals. The observation that there was a constant rate of amino acid substitutions in these genes had been hailed as the "molecular evolutionary clock" – i.e., the rate of change in these sequences could be used as a "clock" to estimate times to most recent common ancestor in species sharing these genes [Zuckerkandl and Pauling, 1965]. Second he extrapolates that rate to the entirety of the genome. Third, he concludes that nucleotide substitution in the history of mammals has been so fast that it is on the order of one nucleotide pair "roughly every 2 yr." The crux of his argument is as follows: this rate is simply too high to be consistent with Haldane's "cost of selection", or, in Kimura's words, "substitutional load".[2] Thus, these changes must be effectively neutral.

Kimura also gave a very elegant mathematical argument, demonstrating that if the neutral theory was true, the rate of change in the genetic constitution of a population should be exactly proportional to the mutation rate, and independent of population size. In a diploid species containing 2N alleles, the probability that an allele will become fixed is 1/2N. If the mutation rate per generation is u, then 2Nu represents the number of new mutants introduced into the population each generation. Thus, if there is no selection, the rate at which new mutations become fixed in a population should be equal to (2Nu)(1/2N), which is equal to u. That is, the rate of change at the molecular level should be constant over time, directly proportional to mutation rate, and independent of population size.

[2]The cost of selection is the selective death that must occur for a gene to be substituted [Haldane, 1957]. For further discussion, see below.

2 RECEPTION OF THE THEORY

When Kimura first proposed his theory, James Crow writes that: "The initial response was generally one of dismay and disbelief. The reactions ranged from skepticism to outright rejection. To some it was utter nonsense." (See [Crow, 1985a, 1].)

On the other hand, Ford Doolittle reports,

> I was a graduate student in the 1960's, but I was a molecular biologist, so the neutral theory struck me as "duh". Because we were used to looking at lots of sequences and saying, "These are the important residues. They must be functional. These others must not matter". So even though I was in Charlie Yanofsky's lab, and he was a panadaptationist – at least at some level — at the same time we could entertain the fact that these amino acid changes that you could see in the spectrum of sequences were neutral. We always said, "What are you guys so excited about, because it's just obvious". (MIT Website in History of Science and Technology, `http://hrst.mit.edu/hrs/evolution/public/transcripts/ideology_transcript.html`)

Thus, there were two extreme responses. On the one hand, some rejected Kimura's arguments. On the other, some regarded his results as obvious. What explains this divide? First, why did the neutral theory strike molecular biologists such as Doolittle as so commonsensical?

Long before the neutral theory was proposed, biochemists and molecular biologists were aware that enzymes – proteins, which are composed of long chains of amino acids with a characteristic shape and function, that assist chemical reactions in the cell – have an "active" site combined with a substrate. The active site is the area of the molecule that contributes to the function of the enzyme in the cell, such as breaking down or transporting other molecules. The functional sites of many important enzymes – e.g. cytochrome c – are uniform among vertebrates. However, what Sueoka [1961] called "dispensible" parts of the molecule, vary across vertebrates. In the 1960s, molecular biologists Margoliash [1963], Zuckerkandl and Pauling [1965] found that the non-functional parts of enzymes evolve in a regular, clock-like fashion, without compromising the enzyme's function in the cell. As early as 1966, Thomas Jukes commented on these observations:

> If we consider as an example the two cytochrome c molecules found respectively in dogs and horses, it will be noted that these differ in about 10 of the amino acids in a chain of 104. The question arises, are the two molecules splendidly tailored to the different requirements specified by "dogfulness" and "horsefulness", have they evolved to conform to these different requirements, or have the two cytochromes been carried along as dogs and horses evolved separately form a common ancestor? ... It is undoubtedly quite probable that separation of the two species

> would be followed by changes in the genes that in time would result in differences in the two cytochrome c molecules. (Jukes, *Molecules and Evolution*: New York: Colombia University Press, 1966, (in [Jukes, 1991, 474]))

So, the fact that many parts of molecules – even some that have a hugely significant function – have "dispensible" as well as functional components, made the neutral theory seem, not only initially plausible to many biochemists, but obvious.

However, the neutral theory was not so obvious to many evolutionists. Even if many changes at the molecular level might be functionally neutral, many biologists were skeptical that all or nearly all genetic variation could have little or no effect on an organism's phenotype. They found the arguments Kimura offered on behalf of this claim unpersuasive.

3 KIMURA'S ARGUMENTS FOR NEUTRAL EVOLUTION

What was the reasoning in favor of this claim? Recall that Kimura claimed that there were too many genetic substitutions for selection to be the primary factor in evolutionary change at the molecular level. He claimed that the "cost of selection" or "mutational load" would be too high. What is the "cost of selection" and why would it be too high? The cost of selection is the selective death that must occur for a gene to be substituted [Haldane, 1957]. Selection causes a certain number of individuals to die in each generation. So, selection imposes a "cost" on the remaining (fitter) members of the population to reproduce at a higher rate to make up the difference. If the remaining members of the population do not reproduce enough to make up the difference, then the population will go into decline. If this goes on long enough, the population could be driven to such low levels that it will become extinct.

As the proportion of individuals carrying the less fit alleles in the population decreases, the cost of selection will also decrease. In other words, as there are fewer and fewer suboptimal individuals in a population, the lower the load on the population will be. So, a "selectionist" – i.e., someone who thinks that most of a population is of high fitness, or very close to an adaptive peak – will think that the load cannot be very high. Many evolutionists took this to imply that if selection was, as Darwin said, constantly "scrutinizing" every trait, then the genetic material should be relatively uniform. The cost of selection will thus place an upper limit on the rate of evolution. The upper limit suggested by Haldane for a diploid population was one gene substitution per 300 generations. Haldane's cost of selection was used to argue that the rates of molecular evolution are too fast to be explained by natural selection.

4 TESTS OF THE NEUTRAL THEORY:

In order to test the neutral theory of molecular evolution, we have to know what it would look like if most of the variation at the molecular level was indeed due to drift, or random fixation of alleles, rather than selection. What would the "signal" of randomness be? Part of the history of the debate over the neutral theory has been over exactly this question; what should we expect if Kimura is correct? Until relatively recently, almost all of the tests of the neutral theory either had little statistical power, (in other words, they could not rule out selection), or were indecisive. Before the 1980's, tests of the neutral theory were conducted viz. examination of protein polymorphisms, or variations detectable with electrophoretic data. In the 1980s and 90s, DNA sequence data became available, and molecular biologists did indeed find a great deal of variation at the genetic level. However, there is variation and there is variation. "Fixed differences" or "divergences" are genetic differences between species; for instance, if one species has nucleotide A at a certain site and another has nucleotide G. In contrast, "polymorphism" is nucleotide variation that distinguishes two alleles within a species. If the neutral theory is true, both types of variation should evolve at the same rate – in other words, the neutral mutation rate should explain both polymorphism within and between species. However, we cannot observe and compare these rates directly; rather, we must examine present patterns of fixed differences and polymorphism and infer backward as to their causes.

Testing the neutral theory is enormously difficult, exactly because there is the difficult problem of differentiating genetic variation that is due to selection from variation that has no effect on fitness. Martin Kreitman nicely sums up the dilemma of searching for signals of selection versus drift as follows:

> The detection of positive selection in DNA sequences poses an immense challenge. The genetic material can be likened to a device that faithfully records every informative event (i.e. mutation) but then over time proceeds to either erase (by back mutation) or obscure (by parallel mutation) some of the recorded information. Furthermore, there is not simply one recorder playing at any one time, but a whole population of them (the gene pool), and each records a slightly different, but correlated, version of history. However, only one of these recordings or, more accurately, a heavily spliced (i.e., recombined) version gets saved for posterity... Which spliced snippets get saved depends upon innumerable chance events, ranging from the relatively benign drift of a neutral mutation in a large population to the strong directional shifts in allele frequencies at sites linked to an adaptive mutation. So, even though every functionally important mutational event in the history of a species, is, by definition, recorded in the DNA sequence of a species, these informative mutations are likely to be embedded in a sea of less meaningful ones (selectively neutral and nearly neutral mutations) and are likely to be associated with stochastic events that can

> result in many possible configurations of linked variation or change. The challenge of detecting selection at the level of DNA is the challenge of finding its signal in a leaky, lossy medium. [Kreitman, 2000, 540]

So, how do biologists meet such a challenge? There are at least twelve different tests of selection at the DNA level (For a Review, see [Kreitman, 2000]). The current state of the field is in flux; new tests are being developed all the time, and claims to have demonstrated the neutral theory over the past twenty years have been closely followed by claims to the contrary. All tests of selection at the DNA level are not tests of the neutral theory, per se. Rather, they assume neutrality, or use the neutral theory as a "null hypothesis", and look for departures from neutrally evolving sites. There have been many cases found of signals of "selection" at the molecular level. However, this does not count as a refutation of the neutral theory. There is no question that some parts of the genome have been strongly constrained by selection, and other parts are not so constrained. In other words, today, the question of whether the neutral theory is correct is not an "either-or" question, (whether there is neutral evolution), but a "more-or-less" question, (that is, what proportion of evolution at the molecular level is neutral).

One example of a very popular test of the neutral theory is the MacDonald-Kreitman, or MK test. The MK test, (or, rather, the family of tests, as there are several versions), compares the ratio of variability in "replacement" and "synonymous" sites. The DNA code is "degenerate", which means that not all changes in the nucleotide sequence entail changes in the amino acid and protein produced by that sequence. Thus, "replacement" sites are sites where changes the nucleotide sequence will change the amino acid sequence of a protein. "Synonymous" sites are sites where changes in the sequence do not change the amino acid sequence of a protein. If the bulk of molecular evolution is neutral, then the variability within a species and the rate of evolution between species are each linearly related to the neutral mutation rate. The MK test compares the number of fixed differences for replacement and silent nucleotide changes, dividing one by another to get a ratio. Then, this is compared with the ratio of polymorphism within species for replacement and silent sites. The two ratios should be equal if all changes are neutral. If the ratios are significantly different, however, then one can predict with some confidence that selection has acted to change the amino acid sequence for some protein.

5 NEUTRALISM & ADAPTIVE EVOLUTION: THE MOLECULAR AND THE PHENOTYPIC LEVEL

What are the implications of these tests? If we discover that many if not all change at the molecular level is neutral, what are the implications for adaptive evolution? Does this mean that many if not most phenotypic traits are not products of selection? The answer is no. It may well be the case that most if not all evolution at

one level is neutral, and that at another level, it is strongly controlled by selection. This may seem like a paradox, but it is not. Phenotypic traits are controlled by many genes, and, there is a good deal of "redundancy" in both the genetic and developmental bases of selectively significant phenotypic traits. In other words, there might be a good deal of "give" between the genetic bases of certain traits and their phenotypic expression. Thus, it may well be the case that most if not all of evolution at the molecular level is due to drift, and at the same time, most if not all evolution at the phenotypic level is a product of natural selection.

Thus, that most traits are controlled by "drift" at the molecular level is not the same as to suggest that random chance and accident has controlled the fixation of most phenotypic traits in evolution. There is a sense in which what happens at the molecular level is relatively independent from that at the phenotypic level. Drift in the classical sense was assumed to operate on discrete genes that controlled discrete phenotypic traits. On the classical model of drift, the main "cause" of drift is change in effective population size; intuitively, this makes sense, as when populations are drastically reduced in size, by chance alone, there will be a radical shift in the distribution of traits in that population.

The classical "Wright-Fisher" model represents drift by random binomial sampling.[3] In other words, we take generations to be discrete, and imagine that alleles (where there are two alleles at a gene locus) are "sampled" from one generation to the next, in the way that balls are drawn from an urn (we are to assume that phenotypic traits are closely associated with specific alleles). So, for instance, consider two individuals in a parent generation, one of which is heterozygote AB and another homozygote, AA. Given the Mendelian assumption of independent assortment, these two individuals can have offspring of one of two sorts, either AA, or AB. By chance alone, they may have an equal number of AA and AB offspring, or, alternatively, ten offspring that are all AA, or ten offspring that are all AB. Summed over the population as a whole, the change in distribution of gene frequencies due to independent assortment is called drift. In other words, the "cause" of a drift in gene frequencies in this sense is simply redistribution due to independent assortment, or accidents of "sampling" of alleles. What is meant by "cause" in this context? Mendelian independent assortment is a cause of drift in

[3]There are several problems with this model of drift. First, Mendelian independent assortment is systematically false. As Sturteyvant pointed out, one can use failure of independent assortment to map locations of genes on chromosomes. Roughly speaking, the less "independent" genes are, the more closely linked they are on a chromosome. Some (Provine, unpubl.) have argued that given the failure of independent assortment, drift in the classical sense is never instantiated. And so, in some sense, there is no such thing as drift, as no populations of organisms in nature are correctly described by the Wright-Fisher model. More worrisome still, Gillespie has pointed out that we may get the same data that we might get were a population to meet these conditions, either by a process of fluctuating selection, or draft. Draft is when genes are swept along to fixation because they are linked to genes that are strongly selected for. They are swept along, as it were, in the "draft" of a selective sweep.

This creates a whole slew of problems for testing drift v. selection. Suffice it to say that it is an idealized assumption of classical models of drift that sampling is "perfectly" random, and that testing claims about the relative significance of drift versus selection in any case is no small feat.

the same way that the fact that automobiles have internal combustion engines is a cause of the long lines for gasoline during the 1973 Middle East oil embargo. It is a condition on the possibility of drift, insofar as drift is understood as binomial random sampling, that there be independent assortment.

It is important to distinguish conditions on the possibility of drift occurring, where drift is represented by some model, from law-like regularities about how the effects of drift can be increased or decreased. For example:

- Decreasing population size increases drift. Or, suppose you have two sets of populations and imagine that there is no selection, mutation, migration or assortative mating in such populations. Time to fixation of alleles in population of smaller size will be much less than in population of larger size. By drift, or chance alone, alleles will drift to fixation more quickly in smaller populations. This is true for the same reasons that one expects a smaller sample of flips of a fair coin to be skewed towards heads or tails than a larger sample.

- Patterns of mating – such as polygamous mating patterns or polyandrous mating patterns – can increase the effects of drift. Such mating patterns reduce what is called the 'effective population size'

- Selfing (or self-fertilization (often occurs in plants)) or assortative mating both increase the effects of drift

- Whenever selection coefficients are smaller than 1/2Ne, drift will win out. Or, the fate of alleles with such selection coefficient will be controlled largely by drift. Selection is operating, but too weak to offset the influences of chance events. "Effective population size" is the size of an idealized population that would have the same effect of random sampling on gene frequency as that in the actual population.

Thus, there are, on the one hand, conditions for the possibility of drift occurring, and on the other, law like general claims about how random fixation of alleles can be speeded up or slowed down.

So, there is a sense in which "chance" in evolutionary biology is subject to laws. This may seem inconsistent, it is not. On the one hand, there are robust, law-like type-level claims about drift. The problem is that we don't always know which of these law-like ways in which the effects of drift (or, chance) can be increased or decreased is in operation. Stochastic gene frequency change will always occur, so long as the conditions mentioned above are met. And, stochasticity can be increased by population bottlenecks, changing mating patterns, or what have you. In this sense, there are law-like explanations of how drift works, or how stochastic changes in gene frequency can be manipulated. There are nomological facts about drift.

6 WHAT IS DRIFT IN THE CONTEXT OF THE NEUTRAL THEORY?

The classical models of drift were generated before biologists knew that genes were composed of DNA. So, what "drifted" were gene frequencies, where it was assumed that inheritance was Mendelian, and there was a direct relationship between genes and phenotypic traits. Of course, neither of these assumptions is strictly true. So, the models and the ways of thinking about the role of chance in changes of phenotypic traits in population has to be adapted to advances in genetics and molecular biology. Models of molecular evolution use the term "drift" to describe fixation due to "chance". However, most of the models of "drift" at the molecular level treat the fixation of genes as a continuous process – specifically, they are diffusion models. In other words, they are continuous approximations of underlying discrete processes. However, the approximations are quite accurate even given this false assumption. The important difference between modeling drift at the molecular level versus the phenotypic level, however, is that, assuming the neutral theory, the rate of fixation of alleles due to drift at the molecular level is independent of population size. This marks a striking difference from classical models of drift, where effective population size is the main predictor of the extent of fixation due to drift. So, what is the "cause" of drift at the molecular level? It turns out that this is a rather difficult question to answer. While it is true that there are systematic regularities or expectations we might have about the extent of polymorphism, or genetic heterozygosity in populations of different sizes, we cannot say that population size exactly is the "cause" of this heterozygosity. Rather, populations of larger size will tend to have a wider or more diverse samples of alleles, and so, alleles will take a longer time to be eliminated or fixed due to random sampling, than will alleles in a smaller sample. But, the causes of these "fixation" events are, by definition, "random" – since, were there to be deterministic regularities in their fixation, they would be due to selection, not drift.

According to Woodward [2003], causal claims are essentially claims about how manipulation of one variable, (or change in the value of that variable) is capable of changing the value of a second variable. Insofar as manipulating population size can increase the effects of drift, one might speak of drift as a "cause" in Woodward's sense of cause [Riesman and Forber, 2005, forthcoming]. However, what is manipulated here is population size, not drift per se. Drift is simply random sampling of alleles from one generation to the next. Reducing population size can increase the effects of random sampling; just as reducing a sample of coin flips can cause a skew of flips toward heads.

However, if Kimura is to be believed, the effects of drift at the molecular level are independent of population size. Moreover, if we take drift to always be in operation when populations are finite, drift will occur with or without manipulation of population size. Insofar as any system which meets conditions on drift described in the above section will be subject to chance, and all populations meet these conditions, it becomes analytically true that whatever change in frequency distribution one observes from one generation to the next is caused, in part, by

drift, whether or not population size has been manipulated. Put differently, any and all change in distribution of gene frequencies in a finite population must be in part due to drift. So, while it is true that changes in frequency distribution may be due to manipulations of population sizes, drift at the molecular level, at least in the case of neutral alleles, is supposed to be independent of such manipulations, and any case of shift in distribution that is not due to deterministic causes, whether population sizes have been manipulated or not, is called "drift". It seems that there are open questions as to what exactly is being "manipulated" in the case of drift, and in what sense drift is a "cause" of evolutionary change.

To sum up: reducing population size increases variance in gene frequency distributions. Woodward has it that whatever manipulation or choice of a value for a variable x that effects a change in some value of a variable y is a cause. So, it seems his model would have it that choosing a small sample is a cause of any resulting skew in distribution of heads or tails. However, not all will agree that this sort of variable causation is the same as "actual" causation. What this debate hinges upon, it seems, is what we are willing count as variables, and whether causes must be relationships between events in space and time. What makes it the case that smoking "causes" cancer? Is it that individual smokers develop cancer, or that there is a population level probability of getting cancer if one is a smoker? What makes it the case that choosing a small sample of coin tosses causes skew in distribution? Does it have to do with the weight of the or the force of the tosses, or, with the fact that populations of tosses are finite and the law of large numbers? It seems that what we want to call the "cause" of a result depends to some extent upon where we want to pitch the explanation. If we're interested in population level changes in distribution, we may refer to small sample size as a "cause". But, this is not to resolve the metaphysical question of where the causes are.

BIBLIOGRAPHY

[Crow, 1985] J. Crow. Motoo Kimura: an appreciation. In T. Ohta and K. Aoki (eds.), *Population Genetics and Molecular Evolution*. Tokyo: Japan Scientific Societies Press, 1985.

[Crow, 1985] J. Crow. Neutrality-selection controversy. In Kimura and Ohta (eds.), *Population Genetics and Molecular Evolution*, pages 1–18. Japan Scientific Societies Press: Tokyo. Springer Verlag: Berlin, 1985.

[Darwin, 1872] C. Darwin. *Origin of Species*. 6^{th} ed. Murray, London, 1872.

[Dietrich, 1994] M. R. Dietrich. The origins of the neutral theory of molecular evolution. *Journal of the History of Biology*, 27(1): 21–59, 1994.

[Dobzhansky, 1955] T. Dobzhansky. A review of some fundamental concepts and problems in populations genetics. *Cold Spring Harbor Symposium in Quantitative Biology*, 20, 1955.

[Gulick, 1888] J. T. Gulick. *J. Linn. Soc. London*, 20: 189–274, 1888.

[Haldane, 1957] J. B. S. Haldane. The cost of natural selection. *J. Genetics*, 55: 511–524, 1957.

[Jukes, 1991] T. H. Jukes. Early Development of the Neutral Theory. *Perspectives in Biology and Medicine*, 14: 473–485, 1991.

[Kimura, 1968] M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217: 624–626, 1968.

[Kimura, 1976] M. Kimura. Discussion forum: molecular evolution. *TIBS*, July, 1976.

[King and Jukes, 1969] J. L. King and T. H. Jukes. Non-Darwinian Evolution. *Science*, 164: 788–798, 1969.

[Kreitman, 2000] M. Kreitman. Methods to detect selection in populations with applications to the human. *Annual Review of Genomics and Human Genetics*, 1: 539–569, 2000. MIT Website in History of Science and Technology:
`http://hrst.mit.edu/hrs/evolution/public/transcripts/ideology_transcript.html`

[Matthen and Ariew, 2002] M. Matthen and A. Ariew. Two ways of thinking about fitness and natural selection. *Journal of Philosophy*, 94: 55–83, 2002.

[Provine, 1988] W. B. Provine. *Sewall Wright and Evolutionary Biology*. University of Chicago Press, 1988.

[Woodward, 2004] J. Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press, 2004.

[Zuckerkandl and Pauling, 1965] E. Zuckerkandl and L. Pauling. Evolutionary divergence and convergence in proteins. In Bryson and Vogel (eds.).*Evolving Genes and Proteins* New York: Academic Press, 1965.

LEVELS OF SELECTION

Robert A. Wilson

1 INTRODUCTION

The generality of the theory of natural selection is both a virtue and, if not quite a vice, at least the source of much spilt ink, spirited discussion, and occasional despair. The theory of selection is general in that the conditions usually held to be necessary for natural selection to operate — that there be heritable variation in the fitness of some trait within a population — is itself neutral about just what entities natural selection operates on. This kind of neutrality has provided evolutionary biologists with much room in which to develop the theory of natural selection, giving rise to "universal Darwinism" [Dawkins, 1983; Dennett, 1995], to both dualistic and integrative accounts of natural and cultural selection [Durham, 1991; Boyd and Richerson; 1985; 2004] and to debate over the levels of selection. It is this final issue that will be our focus in what follows. Sections 2–5 will cover basic issues and views in this debate, while sections 6–8 will take up some more specialized topics that are of current debate: pluralism (section 6), the conception of groups (section 7), and evolutionary transitions (section 8).

If the Darwinian theory of natural selection is neutral in the way the people have typically thought that it is, then there is a real question as to what the agents of natural selection are. What are the *units* on which natural selection acts? Darwin himself formulated his theory of natural selection with *organisms* in mind as the principal (and near exclusive) unit of selection. They were the bearers of the properties amongst which there was variation in a population, the raw material on which natural selection operates, and through their reproductive behavior were the source for the transmission of these traits across generations, and so important entities for inheritance. It is true that in several often quoted passages, Darwin also entertained the idea that *groups of organisms* could also serve these roles, but it is clear that Darwin appealed primarily to *individual selection* and only rarely to *group selection* as the mechanism through which patterns of descent with modification were established in the natural world.

2 FROM ORGANISMS TO GENES AND GROUPS

Such a view of natural selection was largely taken for granted by advocates of Darwin's theory until the Evolutionary Synthesis in the 1930s and 1940s. While

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

the idea that natural selection acts at the level of the individual organism emerged from the Synthesis still a dominant view, during that time developments within the biological sciences led to alternative views that have come to constitute its two chief challengers.

The first of these came from genetics, both in its experimental guise exemplified in "the fly room" established at Columbia University by Thomas Hunt Morgan, and in the development of theoretical population genetics that ranged from the mathematical modeling of Ronald Fisher [1930] and Sewall Wright [1929; 1931; see also 1968; 1969; 1977; 1978; 1980] to the broader applications of Theodosius Dobzhansky [1937]. This work introduced the gene as the principal agent of inheritance: it is genes that are inherited and that are causally responsible for the production of organismic traits. In a metaphor introduced in Erwin Schrödinger's *What is Life?* [1944] that has, since that time, come to dominate the scientific and popular conception of genetic agency, genes *code for* the fundamental products of inheritance. This cluster of developments thus formed the basis for a challenge to the role of the individual organism in the process of inheritance, a crucial part of the overall process of natural selection. The idea that natural selection operates *on genes*, although nascent in remarks in Fisher's *The Genetical Theory of Natural Selection* [1930], was developed in the late 1950s and early 1960s by George Williams [1966] and William D. Hamilton [1964], and received further extension and popularization in Richard Dawkins' first book, *The Selfish Gene* [1976].

The second development within the conglomeration of biological sciences associated with the Evolutionary Synthesis was centred disciplinarily in plant and animal ecology and institutionally at the University of Chicago. The Harvard entomologist William Morton Wheeler had introduced the term "superorganism" in his insightful and playful 1920 essay "The Termitodoxa, or Biology and Society", having earlier suggested conceiving of an insect colony as a higher-order organism in his "The Ant-Colony as an Organism" [1911]. This kind of view of at least some groups of organisms was already implicit in the work of the ecologist Frederic Clements on plant succession, who argued that plant-animal communities formed what he called *biomes*, which were more than simply the sum of the individual organisms that constituted them. Common to both the ecological and entomological works here was a physiological, organismic view of populations that came to occupy centre stage in the Chicago school of ecology, headed by Warder Clyde Allee and whose members included Alfred Emerson and Thomas Park. The conception of at least certain groups of organisms as themselves organism-like served as the basis for viewing such groups as the *beneficiaries* of natural selection, the entities that differentially survive as a result of the action of natural selection, as well as the *manifestors of adaptations*, the entities that come to bear adaptations through that process. This idea of group selection was a thread that ran through the Chicago school's *Principles of Animal Ecology* [1949] and received its now-classic expression in the work of the Scottish ornithologist V.C. Wynne-Edwards, *Animal Dispersion in Relation to Social Behavior* [1962].

So there have been challenges to the traditional Darwinian view of the level at

which selection acts from two fronts: from "the gene below", as well as from "the group above". Over the past 30 years, during which the philosophy of biology has emerged as a distinct sub-discipline within the philosophy of science, this debate has become increasingly sophisticated, in part due to interactions between philosophers and biologists. The question of the level(s) at which selection acts has in fact become a staple in the diet of any healthy philosopher of biologist. It is a question answered in part via an appeal to a range of "isms" — pluralism, realism, reductionism, individualism, genocentrism — and methodologies — the drawing of distinctions (some of which I have already taken advantage of above), the reformulations of questions, the identification and questioning of taken-for-granted assumptions — that are the signature of a philosophical presence.

3 GENES AND ORGANISMS: REPLICATORS, INTERACTORS, AND OTHER "UNITS"

Richard Dawkins has nicely captured one way in which the individual organism acts as a basin of attraction for evolutionary reflection:

> biologists interested in functional explanation usually assume that the appropriate unit for discussion is the individual organism. To us, 'conflict' usually means conflict between organisms, each striving to maximize its own individual 'fitness'. We recognize smaller units such as cells and genes, and larger units such as populations, societies and ecosystems, but there is no doubt that the individual body, as a discrete unit of action, exerts a powerful hold over the minds of zoologists, especially those interested in the adaptive significance of animal behaviour. [1982, 4]

One of Dawkins' own chief aims has been to break the hold that this image of adaptation, function, and natural selection has on biologists. He does this by making genes more focal in the range of evolutionary narratives that we tell about the biological world.

A simple, general distinction, introduced originally by Dawkins in *The Selfish Gene*, has been integral to the intuitive plausibility of the idea that natural selection operates typically at the level of the gene. This is the distinction between *replicators* and *vehicles*. A replicator is anything that can be reliably copied with a high degree of fidelity, while a vehicle is the entity in which a replicator is housed. Dawkins' claim was that genes are the primary replicators, and organisms the chief vehicles, in the process of evolution by natural selection.

The power of this distinction in motivating genic selection is threefold. First, as a general, functional distinction, it can be introduced neutrally and illustrated independently from the debate over the levels of selection. Second, the distinction allows us to probe any putative case of organismic or group selection by asking whether those entities serve as replicators or vehicles. Third, the asymmetry between functionally-active replicators and functionally-passive vehicles suggests

that it is the former which are the real agents or units of selection, and the latter that play some supplementary role in the process of selection.

David Hull [1980] generalized and tidied this distinction — in part to remove the above-mentioned asymmetry between replicators and vehicles — envisaging natural selection as composed of two processes, replication and *interaction*. These may (and Hull thinks, typically do) operate at distinct levels on distinct entities. Following Hull, a replicator is "an entity that passes on its structure directly in replication" and an interactor is "an entity that directly interacts as a cohesive whole with its environment in such a way that replication is differential". Bypassing the apparent circularity in each of these definitions, they open the space for a two-part characterization of natural selection that has become common in the literature (e.g., [Brandon, 1987; Sterelny and Griffiths, 1999]). With these definitions in hand, Hull defines natural selection as "a process in which the differential extinction and proliferation of interactors cause the differential perpetuation of the replicators that produced them" [1980, 318]

Following the standard Darwinian view, organisms are often taken for granted as the principal interactors in natural selection, and this has been central in thinking of them as the agents of selection. Although it is also common to dismiss the idea that sexually reproducing organisms are replicators, this idea deserves closer scrutiny. The reasons for rejecting organisms as replicators include appeals to facts about genetic recombination, the non-identity of parents and offspring at the phenotypic level, and the "indirectness" of the reproduction of organisms through sexual means (where directness is understood in terms of the notion of independence). These, in turn, serve as the basis for arguments that genes are special as agents in natural selection because they are permanent, or the only entities that are inherited across generations. But none of the basic appeals are very good reasons for denying that organisms are replicators, given the characterization of a replicator. Whether organisms are replicators turns largely on at how fine-grained a level we individuate structures, for there is a clear sense in which many phenotypic structures are passed on across generations, even "directly". If we individuate them relatively coarsely, many are transmitted through sexual reproduction and their genetic heritability is relatively high, but that is not strictly required by the definition of a replicator. Philosophers and biologists have been too quick to dismiss the idea that organisms themselves are replicators.

Elisabeth Lloyd [1992; 2001] has distinguished in addition two further issues that are sometimes built into the debate over the levels of selection, what she calls the *beneficiary* and the *manifestor of adaptation* questions. Beneficiaries of selection are those entities that end up being differentially represented in later generations as a result of natural selection. Lloyd plausibly considers species or lineages as putative beneficiaries of selection. Manifestors of adaptation are those entities that come to possess (or even lose) traits as a result of the action of natural selection. Organisms are readily conceived not only as interactors but also as beneficiaries — they are relatively easily counted, for example, and organisms with fitter traits are differentially represented in future generations — and as

manifestors of adaptations — their complex design was obvious before Darwin. Perhaps because of this, these distinct roles are sometimes not distinguished when considering other putative agents of selection, particularly groups. By invoking this additional distinction, Lloyd has turned the tables on the genic selectionist who appeals to the replicator/vehicle distinction to support his view. Just as proponents of traditional individual selection failed to recognize replicators as distinct from vehicles or interactors, so too do genic selectionists lump together genes as replicators, beneficiaries, and manifestors of adaptation in the process of natural selection. On Lloyd's view, there is not one but four questions about the unit of selection, and genic selectionists have concentrated on answering just one of these.

4 GROUP SELECTION AND INDIVIDUAL SELECTION

The most convincing articulation of *group selection* is one that is modeled on and builds on our intuitive views of individual selection. Just as there can be the natural selection of organisms within a population for some fitness-enhancing property — running speed, wing shape, color — so too can there be the natural selection of groups within a population of groups for some fitness-enhancing property. This selection of groups is group selection, just as the selection of individuals is individual selection.

We can build on this intuitive parallel by, in the first instance, being a little more precise about the conditions necessary for natural selection to occur. Following a seminal discussion of the units of selection by Richard Lewontin [1970], natural selection is often thought of as requiring that three things hold of the entities that it acts on. On the standard Darwinian view, these entities are individual organisms, and so on that view there must be:

(i) variation in a population of organisms with respect to some phenotypic trait;

(ii) a correlation between this variation and the fitness levels of organisms within the population; and

(iii) the heritability of this variation across generations.

There are analogs to Lewontin's three conditions for natural selection that take the group rather than the individual organism to be the relevant object of focus. These analogs are that there be:

(i′) variation in a population of *groups* with respect to some phenotypic trait;

(ii′) a correlation between this variation and the fitness levels of *groups* within the population; and

(iii′) the heritability of this variation across generations.

The only difference between (i′)–(iii′) and (i)–(iii) is the occurrence of "groups" in place of "organisms" in the two places emphasized in (i′) and (ii′). The paradigm of laboratory or experimental group selection established by the geneticist Michael Wade [1977; 1978] is based on the satisfaction of (i′)–(iii′), much as the domestic breeding that Darwin appealed to in the opening chapter of the *Origin* was based on the satisfaction of (i)–(iii).

To ensure that this formal parallel is not superficial and misleading, it is crucial that we be able to make sense of notions that occur in each, such as fitness. An individual organism's fitness is its reproductive success, measured ultimately in terms of number of offspring that it produces. One can further distinguish between survivability and fertility as aspects of individual fitness that correspond, respectively, to the egg-adult and adult-egg halves of the life cycle, in order to attain a more fine-grained picture of where individual selection operates in particular cases, although since these have different mathematical representations, dividing fitness up in this way introduces some theoretical inelegance in modeling fitness dynamics. Individuals within the population are said to compete with one another to leave more offspring in future generations, although "competition" is used here in a "large and metaphorical sense", to use Darwin's own characterization of the idea of a struggle for existence. Whenever there is either differential survivability or fertility in a population of organisms, and finite use of resources, then there is competition, in this sense, between individuals in that population.

Group fitness can be understood in much the way that we understand an individual's fitness, namely, in terms of *its* reproductive success. As with individuals, in the case of groups we also have two types of reproductive success that correspond to organismic viability and fertility. A group's viability is its ability to endure over time, and it does this by replacing the individuals in it as individuals die and others are born. A group's fertility is its ability to produce offspring. As with individuals, we might well view fertility as an index of the ultimate evolutionary currency, but in addition, there are two forms that currency can take.

First, groups can produce or recruit *more individuals than they lose*, and so grow in size. Second, groups can produce *more groups* with the characteristics that they have as groups. Both of these involve producing (or recruiting) more individuals over time, but they are in principle independent means of increasing the fitness of the group such that groups with a given individual phenotypic or group trait successfully compete with groups without it, and thus come to replace those groups in the metapopulation. Thus, an individual-level phenotypic trait could increase its representation in the metapopulation via group selection in either of two ways. First, it could do so by the differential addition of individual organisms to existing groups — paradigmatically by one group increasing in size, or a competitor group having its size decreased. Second, it could increase its metapopulational representation by the differential addition of groups of individual organisms with that trait — paradigmatically through differential colonization and migration rates between groups.

As Okasha [2007] has recently pointed out, these two different kinds of group

fitness have been used in the literature in different ways. Proponents of trait group selection, such as Wilson [1975] for example, have been interested in a group-level mechanism that could explain the evolution of an individual-level trait, such as being an altruist. To show that altruism could evolve by group selection, Wilson uses the first measure of fitness, arguing that if groups with more altruists are fitter than groups with fewer altruists, altruism can increase in the global population. Proponents of species selection, such as Jablonski [1986; 1987], have focused on the second measure of fitness. Jablonski argued that species of mollusks with greater geographical dispersal evolved through the group selection of species, and his measure of this was the increased number of species of mollusks with that range, tying this to increased feeding opportunities.

5 THE PROBLEM OF ALTRUISM AND THE LEVELS OF SELECTION

Departures from the standard, individual-centred view of natural selection have their own motivations, as we have seen, but they also share one important motivation: the problem that altruism poses for the standard view. Group selection of the kind originally advocated was one response to the phenomenon of altruism. Genic selection arose as an alternative that was claimed both to avoid the problem that altruism posed for individual selection, and to make appeals to group selection otiose. But what is the problem of altruism?

On the standard Darwinian view, populations of organisms evolve because the individuals in them have differential levels of fitness. As we have seen, those organisms can be said to compete with one another in a large and metaphorical sense for the survival of their offspring. In this same sense, organisms can be thought of as striving to maximize their fitness, i.e., their own survival and ultimately the survival of their progeny. Although organisms are often thought of as striving for their own survival, those that do so to the exclusion of producing viable offspring — say, either by producing no offspring at all or producing none that survive as fertile individuals — have a fitness of zero. Given that the fitness of any given organism is ultimately its expected number of offspring, any individual striving to maximize its fitness will be striving to maximize this number. Thus, it will act in ways that benefit at least some others, i.e., its progeny. But an individual's biological fitness places it in competition with other members of the population, and so individuals who reduce their own fitness in order to increase the fitness of others who are not progeny will reduce their representation in future generations.

Evolutionarily altruistic behavior is typically characterized as behavior that has just this property of reducing an individual's fitness while increasing the fitness of non-offspring in that individual's group. For example, Edward O. Wilson defined altruism as "self-destructive behavior performed for the benefit of others" in the glossary of his influential *Sociobiology: The New Synthesis*. But such behavior is merely an extreme form of a more general type of behavior that gives rise to the problem of altruism. This problem arises just when a behavior contributes relatively more to the fitness of non-offspring in the population than to the fitness

of the individual engaging in the behavior, and thus that decreases the *relative* fitness of the "altruistic" individual within the population. Because individual selection will diminish the relative fitness of individuals engaging in such behaviors from one generation to the next, it will select against them. If unchecked, it will drive them to extinction in the population. It is precisely such behaviors that give rise to the problem of altruism. Thus, these behaviors, which may or may not be "self-destructive" or "performed for the benefit of others", are altruistic, i.e., behaviors for which the problem of altruism arises. Behaviors that are self-sacrificial or that benefit others are merely as a special case. One advantage that this characterization of the problem of altruism has is that it makes it easier to dissociate altruism from self-sacrifice, a notion with a psychological caste that readily comes to mind in thinking of human altruism. The "sacrifice" involved in behaviors for which there is a problem of altruism is just that of the maximization of the number of one's viable offspring.

Given the individual as the agent or unit of selection, the existence of altruistic behaviors, so characterized, would be a puzzle, since individuals in a population who exemplify them will be less fit than those who do not. Thus, other things being equal, such individuals will leave fewer offspring in the next generation than do their competitors. From this perspective, being altruistic is a differential handicap, like being slow relative to others in a population, where greater speed allows one either to capture more prey or to escape more readily from predators. Such fitness-reducing behaviors may be the by-product of selective processes operating on other phenotypes but could not themselves evolve by individual selection.

The problem of altruism, then, is the conjunction of the standard Darwinian view of natural selection with the existence of evolutionary altruism. There are thus two ways to respond to the problem that could be said to represent solutions to the problem, rather than either an admission that the problem reveals the limits of the theory of natural selection (defeatism), or a denial that there is a problem at all for the standard Darwinian view to face (blind optimism).

The first is to deny the existence of evolutionary altruism. Given a range of often-cited cases — for example, sentinels in birds, caste specialization in social insects, "good Samaritan" behavior in humans — in which individuals help others or even sacrifice their lives for others — such a denial might be thought to lack credibility as a response to the problem of altruism. However, altruistic behavior is not simply helping or sacrificial behavior, but behavior that detracts from the relative fitness of the individual. So to demonstrate the existence of evolutionary altruism one cannot simply point to clear instances in which individuals help others or sacrifice themselves for the sake of others. For such behaviors might themselves be a way of maximizing individual fitness. This is the idea behind reciprocal altruism [Trivers, 1971]: individual's forego or limit their own direct reproductive opportunities in order to maximize their long-term fitness through gaining reciprocal benefits from those they benefit. Here individuals are still maximizing their own fitness, albeit indirectly. Hence these behaviors only appear to be evolutionarily altruistic. In effect, this response plays up the role of individual

fitness within evolutionary theory so that there is little or no room for evolutionary altruism. For it to solve the problem of altruism the net benefits to individuals engaged in "altruism" must be greater than the net benefits to those they help.

The second is to modify the standard Darwinian view so as to posit some other unit of selection, and then show how selection operating at that level could give rise to evolutionary altruism. Thus, proponents of group selection have pointed out that although individual selection acts so as to decrease the representation of altruists within a population, groups of altruistic individuals may have a higher level of fitness than non-altruistic groups [Sober and Wilson, 1994]. It follows that a process of group selection will act in a countervailing direction to that of individual selection, and thus altruists could survive as members of fitter groups. This version of the second response goes hand-in-hand with the idea that the traditional Darwinian view requires augmentation, and that there is a plurality of levels at which natural selection operates.

An alternative way to depart from the traditional Darwinian view is more radical in that it involves recasting the theory of natural selection (and thus fitness) in terms of the survival not of organisms but of the genes they contain. If genes are the agents of selection, then organisms can be altruistic if their behaviors maximize the fitness of genes that happen to be located within those organisms. Since not just progeny of a given organism but individuals related in other ways to it, such as siblings and cousins, bear a genetic relationship to that organism, altruism directed at those individuals may be a way of maximizing the fitness of that organism's genes. This is a common way of understanding Hamilton's [1964] kin selection theory (but see [Wilson and Sober, 1998, 66–67]). In effect, this view also denies the existence of evolutionary altruism, and thus implies that both conjuncts that constitute the problem of altruism are false.

There is an important asymmetry between genic and group selection that can be made more explicit by posing two questions:

(a) Does the traditional Darwinian view provide us with a complete or exhaustive view of evolution by natural selection?, i.e., are there evolutionary phenomena that this conception of the agent of selection leaves out?

(b) More radically, are the appearances here actually misleading?, i.e., are there other agents that are in general better candidates for the agent of selection than the organism?

Proponents of genic selection answer "Yes" to (b) because they think that genes are better candidates than organisms for the agent of selection. In part, this is because the gene's eye view of evolution provides a solution to the problem of altruism. Proponents of group selection, by contrast, answer "No" to (a) because they think that certain phenomena (e.g., altruism) require group selection. Thus, they hold that such a process must be added to individual selection to understand the complexity to the biological world. In fact, proponents of group selection are typically happy enough to embrace levels of selection smaller than the organism,

such as the gene, as part of an overarching *multilevel* approach to understanding natural selection [Wilson and Sober, 1994; Sober and Wilson, 1998].

6 PLURALISM AND REALISM

This way of thinking of the relationship between the problem of altruism and the debate over the levels of selection suggests the view, widely accepted in the literature, that there is a fact of the matter about what *the* unit of selection is in at least some particular cases. If altruism really did evolve as a result of the action of natural selection, and the individual organism cannot be the unit on which it acted to produce that result, then genic and group selection are alternative accounts of how altruism evolved, alternatives that paint different pictures of biological reality. But some have denied that there is *always* a fact of the matter about what *the* unit of selection is, while others have claimed that there is a sense in which there is *never* (or rarely) such a fact of the matter. This dialectic raises issues of pluralism, realism, and reductionism into the debate over the levels of selection.

Realists about the level of selection hold that there is a fact of the matter, either in particular cases or in general, about what level selection operates at. They are like realists in other areas of science, holding that even if we have only limited evidence about what the underlying processes are that generate the phenomena we observe and measure, nonetheless there is an "in principle" fact of the matter here. The properties of electrons could not be reliably measured until into the 1920s, and the biochemical structure of genes was not known until the 1950s, but there was a fact regarding each of these matters that awaited our discovery. Likewise, even if it is hard for us to tell about whether, say, altruism evolves by genic or by group selection, either in a particular case or more generally, there is nonetheless a fact of the matter here. Those who deny this are *anti-realists* about the levels of selection.

The issue of realism is sometimes run together with, but is actually orthogonal to, the question of whether there is *a* level of selection, or whether there are multiple levels of selection. *Monists* hold the view that there is a single level of selection, either in a particular case or more generally, while *pluralists* deny this, holding that natural selection can and indeed does operate at different levels. In the case of altruism, a monist must view genic and group selection as exclusive alternatives to one another, such that at most one of these processes is causally responsible for the evolution of altruism. A pluralist, by contrast, could allow either that both genic and group selection act in a particular case to promote altruism, or that genic selection operates in some cases, group selection in other cases. Either way, there is no *single* level of selection that is responsible for the evolution of altruism, but a plurality of levels. According to pluralists, to ask the question "What is *the* level at which selection operates?" is to make a mistake insofar as this question invites monistic responses.

It is very natural to elide this pair of distinctions and (roughly speaking) equate monism with realism and pluralism with anti-realism, something encouraged by

the kind of sloganeering that has crept into the study of science. Monists/realists think there is One True Description of the world, a God's-eye view, whether or not we mere mortals can arrive at it, while pluralists/anti-realists think that "anything goes", that reality is socially constructed rather than discovered, that it's "different strokes for different folks" when it comes to our view of ontology. Although this is a mistake in general, I'll concern myself here with the kind of mistake it is when applied to debate over the levels of selection. Not only would such a collapse of the realism vs anti-realism and monism vs pluralism distinctions gloss over a conceptual distinction, but it would serve only as a misleading caricature of the kinds of realism that are monistic, and the kinds of pluralism that are anti-realist. In addition, it would leave no room to characterize positions about the levels of selection that have been articulated and defended by some of the leading proponents in the field. In particular, there are varieties of positions that are properly characterized as *pluralistic realism*, and others that can be viewed as forms of *monistic anti-realism*. I will concentrate on pluralistic realism, which has gained much currency in recent years amongst both philosophers and biologists, but begin with monistic forms of realism.

The traditional Darwinian view is, for the most part, monistic about the level at which natural selection operates: the individual is the unit of selection (with the exception of occasional bouts of selection on tribes or groups). Genic selection has also often been viewed as a form of monistic realism in that it claims that it is the gene, rather than the individual organism, that is the *real* level at which natural selection occurs. This is a kind *genic fundamentalism* insofar as it views the gene as the fundamental unit of selection. On this view, genic selection might correlate with individual selection in a range of cases, and so models cast in terms of individual selection give the right answer in response to the question of what evolves in a particular case. But it is genic selection that constitutes the mechanism generating the distribution of traits that evolves. It is for this reason that genic selection is sometimes viewed as a *reductionist* view of the levels of selection.

There are at least two kinds of pluralistic realism, one of which retains strains of monism. The first is exemplified by the work of Elliott Sober and David Sloan Wilson, especially their *Unto Others* [1998]. In a series of publications they have articulated and defended a *multilevel view of selection*, which, as the name implies is a form of pluralism. On this view, one can begin with a question about the effects of natural selection at a given level (say, that of the organism), but then *frameshift* up (to groups) or down (to genes) to reformulate that very question. This multilevel view is a form of realism about the levels of selection insofar as it implies that there is a fact of the matter that determines the answer to each of these questions asked. In some cases, natural selection will operate at just a single level, at others it will operate at a different level, and in yet others it will operate at multiple levels at once. The last of these cases is most clearly pluralistic, but pluralism is also manifest in the multilevel framework by reflecting on the preceding two cases *together*, for together they imply that there is no overall, single level at which selection operates. The strain of monism in the multilevel view comes out

by considering cases in which just a single unit of selection is involved: although there is an overall pluralism, since this unit can vary across different cases, there is a *local monism*. I have elsewhere called this form of pluralism *unit pluralism* [R.A. Wilson, 2003] and *agent pluralism* [R.A. Wilson, 2005], since it is pluralistic about the units or agents of selection themselves.

A second form of pluralistic realism abandons monism altogether, holding that a plurality of units of selection is *always* present when natural selection operates. This is a view that is relatively undeveloped in the literature but one with which I have considerable sympathy. It is motivated by two considerations. The first is a view of the biological world as inherently complex, variable, and diverse. Even the oldest and, by some lights, the simplest organisms have many specialized parts, including parts that are specialized for replication and reproduction, and they nearly always interact with the rest of the world along with conspecifics and other group-mates. The second is a view of our categories and models for explaining this complexity, variability, and diversity: they are meager, simplifications that allow us to make certain kinds of predictions, but that never do full justice to the raw phenomena. Thus, we conceptualize the biological world in terms of distinct "levels", model how entities at each of those levels behave under certain conditions, and arrive at monistic or pluralistic views of the levels of selection. But this notion of levels is a kind of metaphor, one that carries with it limits and biases, and I have suggested that a metaphor that conforms better to the first point might be that of *entwinement* or *fusion* [R.A. Wilson, 2003]. This form of pluralistic realism posits a significant mismatch between biological reality and our epistemic grip on that reality. Such a mismatch warrants viewing the monistic strain in the multilevel selection view as a reflection of our ignorance, rather than of the biological world itself. Thus, monism itself is the result of a kind of simplification of an inherently messy biological reality that metaphors like that of "levels" fail to do justice to.

A distinct form of pluralism from both of these is what I have elsewhere [R.A. Wilson, 2003; 2005, ch.10] called *model pluralism*, since it adopts a pluralistic view of our models of the biological world. This form of pluralism holds that various *prima facie* distinct models of natural selection, such as selfish gene theory and group selection theory, are actually non-competing accounts of one and the same reality. Model pluralists maintain that although there may be strategic or pragmatic advantage to using one rather than another model in a particular case, these models do not compete for, or share, the truth about the nature of natural selection.

Model pluralism has gained much currency in recent debates over the levels of selection. It has been defended by biologists, such as Lee Dugatkin and Hudson K. Reeve [1994) and Andrew Bourke and Nigel Franks [1995], and receives its crispest expression in the recent work of Benjamin Kerr and Peter Godfrey-Smith [2002]. Sober and Wilson have also embraced model pluralism, saying that inclusive fitness theory, selfish gene theory, and the theory of group selection that they propose are part of a "happy pluralistic family" of alternative perspectives on natural selection that are "simply different ways of looking at the same world." [1998, 98]. When

it is put in this way, model pluralism seems to imply a form of anti-realism about the levels of selection.

Model pluralism is motivated in part by the sense that participants in the debate over the levels of selection are "talking past one another", or that the debate is "just semantic". A core part of the support for model pluralism is the idea that there is an important sense in which the various models of natural selection *are equivalent* so that although there is a sense in which they carve up the world differently, this difference is underpinned by deeper affinity that these views share. Kerr and Godfrey-Smith, in particular, have shown how to translate key terms and equations between what they call the *individualist or contextual* models and what they call the *multilevel or collective* models. If these models are equivalent, then, model pluralists argue, it makes no sense to argue for one of the models rather than another as telling us how natural selection works, either in particular or in general. There are differences between these models, to be sure, but the choices between them are to be made on pragmatic grounds, rather than on the basis that one gets at what's really happening, while the other doesn't.

7 GROUPS AS CONTEXTS, GROUPS AS SUPERORGANISMS

Part of what separates proponents of genic and individual selection, on the one hand, from advocates of group selection, on the other, is their respective conceptions of the place of the structures of populations in the theory of evolution. We can distinguish three conceptions, each corresponding to a distinctive view of what appeals to group selection amount to.

The first is the conception of a group implicit in the early work on group selection associated with the Chicago school of ecology and people such as V.C. Wynne-Edwards. As the preceding discussion suggests, the conception of a group was very much that of an organism-like entity; groups, at least some of them, were *superorganisms* or sufficiently like organisms in the relevant respects to warrant treating them as organisms, and to treat their individual members as parts of that organism. That is why it seemed relatively unproblematic to shift the unit of selection from the individual to the group, since in effect this was simply to apply it to a different kind of individual.

The problem with this conception of a group, as Sterelny [1996] and others have pointed out, is that there are very few groups of organisms that can properly be viewed in this way: they are chiefly found in the social insects, which have a reproductive division of labour and feature sterile castes that don't reproduce at all or do so only under highly restricted conditions. Despite this *prima facie* problem, this conception of a group persists in several contemporary discussions. For example, Sterelny himself concedes that group selection can occur when there are superorganisms, and Wilson and Sober's multilevel view of selection, especially the version formulated in their [1994], seem to say much the same thing in employing their frameshifting model. Frameshifts "up" to the group, as well as "down" to the gene, are justified just when those entities *function as individuals*. It is just

that, unlike Sterelny, Sober and Wilson think that this is rather often the case.

Quite a different conception of groups lies at the core of the revival of group selection, however. This conception is of a group as a "sphere of influence", an aggregate of organisms that "share a common fate". Both of these phrases have been used by David Sloan Wilson [1975; 1983; 1997; 2002] and by Sober and Wilson [1994; 1998] in characterizing *trait groups*. On this view, a group is any aggregate of organisms whose evolutionary fates are tied together, no matter how temporarily or for what reason. In a striking example, Sober and Wilson consider a pair of crickets that find themselves sharing a leaf to cross a river as an example of a trait group, for whatever evolutionary outcome greets one will also greet the other.

Clearly, trait groups do not face the problem of rare instantiation that groups as superorganisms do; in fact, they might be thought to suffer from just the opposite problem, that of being a little too common. Consider clones of the myxoma virus located on a rabbit. It might be thought that all of those clones form a trait group, for if the rabbit dies they all die (they can't survive just on the carcass, or not for long). Yet clones that are located on lesions also form a trait group, since they have the same probability of being transmitted further by mosquito or flea vectors, which are attracted to lesions on an already infected rabbit. And since rabbits can transmit the virus to one another through direct contamination, clones located on rabbits within a hutch also constitute a trait group. This suggests that if there is any group selection in this example — an example well-known in the levels of selection literature in part because Richard Lewontin, in an influential paper [1970], claimed that it was an example whose details could *only* be explained by an appeal to group selection, a claim endorsed by Sober and Wilson [1998], amongst others — there is lots of it. That may be the right thing to say, but if so this in turn implies that there are many, many "levels" at which selection operates in at least a wide range of cases, and, to return to an earlier theme, gives reason to question how accurately the metaphor of levels captures biological reality (see [R.A. Wilson, 2004]).

Both of these conceptions of groups make groups agents or units of selection. Those skeptical of how widespread group selection is in nature, particularly proponents of genic and individual selection, acknowledge that groups are often important in natural selection but claim that both traditional and neo-group selectionists are mistaken about the role that groups play. Rather than being agents of selection, groups serve as a part of the *context* in which genic and individual selection acts. In the myxoma case, natural selection is taken to select *individual clones* but to do so relative to the group environment they are in. Those on a given rabbit face common selection pressures, but they do so individually. Selection is always sensitive to an individual's environment: a dark moth will be selected for over its lighter variants in an environment in which the trees have been blackened through industrialization, but not in environments in which the bark of trees remains light-coloured. The insight of group selectionists, it is claimed, is to draw explicit attention to *population structure* as a significant part of an organism's

(or a gene's) environment. But this doesn't mean that groups themselves are agents of selection, any more than it means that the dark-coloured sub-group of the population of moths is selected through group selection.

This conception of groups as the contexts in which other kinds of entity, such as individuals and genes, are selected, plays a key role in model pluralism. For this forms one half of the two kinds of view — what Kerr and Godfrey-Smith call the *individualist/contextual* view as opposed to the *multilevel/collective* view of natural selection — that they claim are, in some sense, equivalent. But it also plays a role in conceptions of genic selection that accord the gene a fundamental status as a unit of selection, since it reinforces the divide introduced by the distinction between replicators and vehicles: genes have properties that make them suitable as agents of selection, while individuals and groups, even if occasionally manifesting such properties, are more usually conceptualized as playing a background, supplementary role to genic action.

A common complaint lodged against the contextualist conceptualization of groups [Brandon, 1987; Sober, 1984; Lloyd, 2001; 2005] is that it fails to reveal the causal dynamics that appeals to superorganisms and trait groups reveal. As it is sometimes put, accounts that focus on genes are accurate "bookkeeping devices" for what evolutionary outcomes develop, but they often do not reveal the *mechanisms* through which those outcomes are achieved. By providing detailed causal representations of both the contextualist and collective views of natural selection, Kerr and Godfrey-Smith [2002] have done much to reply to this kind of objection, though whether it is enough remains an issue subject to further debate [R.A. Wilson, 2003; Lloyd, 2005].

8 TRANSITIONS IN EVOLUTIONARY HISTORY

The traditional Darwinian view of the levels of selection also gives rise to another question, one that has been informed by recent work on what John Maynard Smith and Eörs Szathmary [1995] have called the *major transitions* in evolution. The kind of individual organisms that we know and love best — multicellular, eukaryotic beasts like ourselves — emerged sometime during the 3.8–4 billion years during which there has been life on Earth. They are themselves a product of evolutionary processes, a kind of evolutionary achievement that represented a new type of biological organization. It is plausible, for many reasons, to think that multicellular organisms evolved from unicellular organisms, and that organisms with eukaryotic cells evolved from organisms with prokaryotic cells, with each of these evolutionary shifts constituting a major transition in evolution. More generally, one can consider the history of life as comprised of a series of such major transitions, starting with the simplest forms of living things at the origin of life, and ending (for now, at least) with the diverse collection of organisms that we see today. Many of today's organisms have many complex parts (including parts there were likely independently living organisms in the evolutionary past, such as mitochondria in animal cells and chloroplasts in plant cells), many form

integrated conspecific and inter-specific groups, and some manifest traits that we usually consider to be social or cultural.

If the major transitions in the history of life are products of natural selection, then there is a obvious way in which the debate over the levels of selection is relevant to thinking about those transitions. For consider, say, the transition from unicellular to multicellular life, something that happened somewhere between one and two billion years ago. Since multicellular life is the product of this transition, natural selection cannot operate on *it*. Thus, if that process was one of individual selection, it must have been one that applied to existing kinds of organisms: unicellular organisms. But we can then raise just these same considerations about the relatively complex unicellular organisms that were the direct ancestors of the first multicellular organisms to arrive at the idea that the very first organisms must themselves have evolved *from something else*. If this is right, then the organism itself is an evolutionary achievement, and so natural selection cannot always have acted at the organismic level. The most plausible candidate for the earliest unit of selection is something like a simplified gene, a self-replicating sequence of DNA (or perhaps RNA), selection on which (somehow) gave rise to the first organism.

This perspective on the major transitions of evolution has been taken up by Richard Michod [1999] and Samir Okasha [2007]. Here the debate over the levels of selection has not simply been applied to shed light on how to think about the major transitions; an understanding of the major transitions has also been viewed as offering support for specific views within that debate. In particular, much of the literature that conjoins these two issues adopts the view that at least some of the major transitions of evolution can be made sense of only (or best) by positing a process of genic (or gene-like) selection. This supports a version of the view that Dawkins expressed in several well-known passages in *The Selfish Gene*, which claimed that genes "ganged up" for form groups and (finally) us: we are the "lumbering robots" that are the evolutionary result of selfish genes acting in ways to further their own interests.

Although there is much here that is interesting, and some points that are clearly correct, I am more skeptical about some of the inferences drawn here and what is apparently presumed. Part of the reason for being cautious is the relative paucity of hard evidence to support a specific series of major transitions, and so for the necessarily speculative nature of many of the relevant empirical claims. But some of the caution issues from more purely philosophical, conceptual concerns. For example, although it is clear that multicellular, eukaryotic organisms are a relative recent evolutionary innovation, it is much less clear that the same is true of *organisms* per se. Minimally, the claim that genes or gene-like entities preceded organisms in the history of life, and so were the original or at least early units of selection, turns in part on what we think organisms are. My own view is that on the most plausible conception of an organism, according to which an organism is a living thing that forms part of a replicative lineage and has some kind of internal control and external freedom — what I have elsewhere called the *tripartite view of organisms* (see [Wilson, 2005, ch.3]) — organisms likely appear very early in the

history of life, certainly within the first billion years of that history.

Likewise, the idea that Dawkins expresses mawkishly through his talk of "ganging up" and "lumbering robots" requires closer scrutiny than it has received to date. Suppose that we waive the preceding issue and grant that selfish replicators of some kind predate organisms. Since there is more than just one such replicator — in fact, the idea is that there are many of them — there is at least the possibility that successful replicators are subject to natural selection *as a group*, rather than individually. And since such replicators likely have some kind of internal complexity, with internal, specialized parts that perform specific functions, it is also possible that natural selection could operate on *those parts*, and so only derivatively on those replicators as entities that have those parts. Thus, we seem to have just the kind of hierarchy of levels on which selection might operate that we have in the contemporary debate over the levels of selection, except with selfish replicators taking the place of organisms.

BIBLIOGRAPHY

[Alexander and Borgia, 1978] R. Alexander and G. Borgia. Group Selection, Altruism, and the Levels of Organization of Life, *Annual Review of Ecological Systematics* 9: 449-474, 1978.

[Allee *et al.*, 1949] W. C. Allee, A.E. Emerson, O. Park, T. Park, and K.P. Schmidt. *Principles of Animal Ecology*. Philadelphia, PA: W.B. Saunders, 1949.

[Bourke and Franks, 1995] A. F. G. Bourke and N.R. Franks. *Social Evolution in Ants*. Princeton, NJ: Princeton University Press, 1995.

[Boyd and Richerson, 1985] R. Boyd and P. Richerson. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press, 1985.

[Boyd and Richerson, 2004] R. Boyd and P. Richerson. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press, 2004.

[Brandon, 1987] R. N. Brandon. The Levels of Selection: A Hierarchy of Interactors. In H. Plotkin (editor), *The Role of Behavior in Evolution*. Cambridge, MA: MIT Press, 1987.

[Brandon and Burian, 1984] R. N. Brandon and R.M. Burian, (editors). *Genes, Organisms, Populations: Controversies Over the Units of Selection*. Cambridge, MA: MIT Press, 1984.

[Buss, 1987] L. Buss. *The Evolution of Individuality*. Princeton, NJ: Princeton University Press, 1987.

[Crozier and Pamilo, 1996] R. H. Crozier and P. Pamilo. *Evolution of Social Insect Colonies: Sex Allocation and Kin Selection*. New York: Oxford University Press, 1996.

[Damuth and Heisler, 1988] J. Damuth and I.L. Heisler. Alternative Formulations of Multilevel Selection, *Biology and Philosophy* 3: 407-430, 1988.

[Darwin, 1964] C. Darwin. *On the Origin of Species: A facsimile of the first edition*. Cambridge, MA: Harvard University Press, 1964. 1^{st} edition, 1859.

[Darwin, 1981] C. Darwin. *The Descent of Man, and Selection in Relation to Sex*. Princeton, NJ: Princeton University Press, 1981. 1^{st} edition, 1871.

[Dawkins, 1982] R. Dawkins. *The Extended Phenotype*. Oxford: Oxford University Press, 1982.

[Dawkins, 1983] R. Dawkins. Universal Darwinism. In B.S. Bendall (editor), *Evolution from Molecules to Man*. Cambridge: Cambridge University Press, 1983.

[Dawkins, 1989a] R. Dawkins. *The Selfish Gene*, 2^{nd} ed. New York: Oxford University Press, 1989. 1^{st} edition, 1976.

[Dawkins, 1989b] R. Dawkins. The Evolution of Evolvability. In C. Langton (editor), *Artificial Life*. Sante Fe, NM: Addison-Wesley, 1989.

[Dennett, 1995] D. C. Dennett. *Darwin's Dangerous Idea*. New York: Simon and Schuster, 1995.

[Dobzhansky, 1937] T. Dobzhansky. *Genetics and the Origins of Species*. New York: Columbia University Press, 1937.

[Dobzhansky, 1973] T. Dobzhansky. Nothing in Biology Makes Sense Except in the Light of Evolution, *American Biology Teacher* 35: 125-129, 1973.

[Dugatkin, 2002] L. A. Dugatkin. Will Peace Follow? *Biology and Philosophy* 17: 519-522, 2002.
[Dugatkin and Reeve, 1994] L. A. Dugatkin and H.K. Reeve. Behavioral Ecology and Levels of Selection: Dissolving the Group Selection Controversy. In P.J.B. Slater et al. (editors), *Advances in the Study of Behavior*, vol.23. New York: Academic Press, 1994.
[Dupré, 1993] J. Dupré. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press, 1993.
[Durham, 1991] W. H. Durham. *Coevolution: Genes, Culture, and Human Diversity*. Palo Alto, CA: Stanford University Press, 1991.
[Eldredge, 1985] N. Eldredge. *Unfinished Synthesis*. New York: Oxford University Press, 1985.
[Eldredge and Salthe, 1984] N. Eldredge and S. Salthe. Hierarchy and Evolution, *Oxford Surveys in Evolutionary Biology* 1: 182-206, 1984.
[Emerson, 1939] A. E. Emerson. Social Coordination and the Superorganism, *American Midland Naturalist,* 21: 182-209, 1939.
[Emerson, 1942] A. E. Emerson. Basic Comparisons of Human and Insect Societies, *Biological Symposia,* 8: 163-176, 1942.
[Emerson, 1943] A. E. Emerson. Ecology, Evolution and Society, *American Naturalist*, 77: 97-118, 1943.
[Emerson, 1946] A. E. Emerson. The Biological Basis of Social Cooperation, *Illinois Academy of Science Transactions,* 29:9-18, 1946.
[Ereshefsky, 1992] M. Ereshefsky, (editor). *The Units of Evolution: Essays on the Nature of Species*. Cambridge, MA: MIT Press, 1992.
[Ewald, 2002] P. Ewald. *Plague Time: The New Germ Theory of Disease*. New York: Anchor Books, 2002.
[Ewald, 1993] P. W. Ewald. *The Evolution of Infectious Disease*. New York: Oxford University Press, 1993.
[Fenner and Fantini, 1999] F. Fenner and B. Fantini. *Biological Control of Vertebrate Pests: The History of Myxomatosis–an Experiment in Evolution*. Oxford: CABI Publishing, 1999.
[Fenner and Ratcliffe, 1965] F. Fenner and F.N. Ratcliffe. *Myxomatosis*. Cambridge: Cambridge University Press, 1965.
[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930.
[Frank, 1997] S. A. Frank. The Price Equation, Fisher's Fundamental Theorem, Kin Selection, and Causal Analysis, *Evolution* 51: 1712-1729, 1997.
[Frank, 1998] S. A. Frank. *Foundations of Social Evolution*. Princeton, NJ: Princeton University Press, 1998.
[Futuyma, 1979] D. J. Futuyma. *Evolutionary Biology*. Sunderland, MA: Sinauer Associates, 1979.
[Godfrey-Smith and Kerr, 2002] P. Godfrey-Smith and B. Kerr. Group Fitness and Multi-Level Selection: Replies to Commentaries, *Biology and Philosophy* 17: 539-549, 2002.
[Goodnight, 1985] C. Goodnight. The Influence of Environmental Variation on Group and Individual Selection in a Cress, *Evolution* 39: 545-558, 1985.
[Goodnight and Stevens, 1997] C. Goodnight and L. Stevens. Experimental Studies of Group Selection: What Do They Tell Us About Group Selection in Nature?, *American Naturalist* 150 (supp.): S59-S79, 1997.
[Gould, 2001] S. J. Gould. The Evolutionary Definition of Selective Agency, Validation of the Theory of Hierarchical Selection, and Fallacy of the Selfish Gene. In R.S. Singh, C.B. Krimbas, D.B. Paul, and J. Beatty (editors), *Thinking About Evolution: Historical, Philosophical, and Political Perspectives*. New York: Cambridge University Press, 2001.
[Gould, 2002] S. J. Gould. *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press, 2002.
[Grafen, 1984] A. Grafen. Natural Selection, Kin Selection, and Group Selection. In J. Krebs and N. Davies (editors), *Behavioural Ecology: An Evolutionary Approach*. London: Blackwell, 1984.
[Grantham, 1995] T. A. Grantham. Hierarchical Approaches to Macroevolution: Recent Work on Species Selection and the 'Effect Hypothesis', *Annual Review of Ecology and Systematics* 26: 301-321, 1995.
[Griffiths, 2002] P. E. Griffiths. Review of Sober and Wilson 1998, *Mind* 111: 178-182, 2002.
[Haig, 1993] D. Haig. Genetic Conflicts of Human Pregnancy, *Quarterly Review of Biology* 68: 495-532, 1993.

[Hamilton, 1963] W. D. Hamilton. The Evolution of Altruistic Behavior, *American Naturalist*, 97:354-356, 1963.

[Hamilton, 1964] W. D. Hamilton. The Genetical Evolution of Social Behaviour I, *Journal of Theoretical Biology* 7: 1-16. Reprinted in his *Narrow Roads of Gene Land*, 1964.

[Hamilton, 1964] W. D. Hamilton. The Genetical Evolution of Social Behaviour II, *Journal of Theoretical Biology* 7: 17-52, 1964. Reprinted in his *Narrow Roads of Gene Land*.

[Hamilton, 1967] W. D. Hamilton. Extraordinary Sex Ratios, *Science* 156: 477-488. 1967. Reprinted in his *Narrow Roads of Gene Land*.

[Hamilton, 1996] W. D. Hamilton. *Narrow Roads of Gene Land, Volume 1: Evolution of Social Behavior*. New York: Freeman, 1996.

[Heisler and Damuth, 1987] I. L. Heisler and J. Damuth. A Method for Analyzing Selection in Hierarchically Structured Populations, *American Naturalist* 130: 582-602, 1987.

[Hull, 1980] D. Hull. Individuality and Selection, *Annual Review of Ecology and Systematics* 11: 311-332, 1980.

[Hull, 1981] D. Hull. Units of Evolution: A Metaphysical Essay. In U.L. Jensen and R. Harré (editors), *The Philosophy of Evolution*. Brighton, Sussex: Harvester Press, 1981.

[Hull and Ruse, 1998] D. Hull and M. Ruse (editors). *The Philosophy of Biology*. New York: Oxford University Press, 1998.

[Jablonski, 1986] D. Jablonski. Larval Ecology and Macroevolution in Marine Invertebrates, *Bulletin of Marine Sciences* 39: 565-587, 1986.

[Jablonski, 1987] D. Jablonski. Heritability at the Species Level: Analysis of Geographic Ranges of Cretaceous Mollusks, *Science* 238: 360-363, 1987.

[Kerr and Godfrey-Smith, 2002a] B. Kerr and P. Godfrey-Smith. Individualist and Multi-Level Perspectives on Selection in Structured Populations, *Biology and Philosophy* 17: 477-517, 2002.

[Kerr and Godfrey-Smith, 2002b] B. Kerr and P. Godfrey-Smith. On Price's Equation and Average Fitness, *Biology and Philosophy* 17: 551-565, 2002.

[Kitcher, 2003] P. Kitcher. *In Mendel's Mirror: Philosophical Reflections on Biology*. New York: Oxford University Press, 2003.

[Lack, 1954] D. Lack. *The Natural Regulation of Animal Numbers*. New York: Oxford University Press, 1954.

[Lewontin, 1970] R. Lewontin. The Units of Selection, *Annual Review of Ecology and Systematics* 1: 1-18, 1970.

[Lewontin, 1983] R. Lewontin. Gene, Organism, and Environment. In D.S. Bendall (editor), *Evolution from Molecules to Men*. Cambridge: Cambridge University Press, pp.273-285, 1983.

[Lewontin, 2000] R. Lewontin. *The Triple Helix: Gene, Organism, and Environment*. Cambridge, MA: Harvard University Press, 2000.

[Lloyd, 1992] E. A. Lloyd. Unit of Selection, in E. Fox Keller and E.A. Lloyd (editors), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, 1992.

[Lloyd, 1994] E. A. Lloyd. *The Structure and Confirmation of Evolutionary Theory*. Princeton, NJ: Princeton University Press, 1994. Reprint of 1988 edition, Greenwood Press, with a new preface.

[Lloyd, 2005] E. A. Lloyd. Units and Levels of Selection, in the *Stanford Encyclopedia of Philosophy*, 2005. Online @: http://plato.stanford.edu/entries/selection-units/

[Lloyd, 2001] E. A. Lloyd. Units and Levels of Selection: An Anatomy of the Units of Selection Debates. In R.S. Singh, C.B. Krimbas, D.B. Paul, and J. Beatty (editors), *Thinking About Evolution: Historical, Philosophy, and Political Perspectives*. New York: Cambridge University Press, 2001.

[Maynard Smith, 1976] J. Maynard Smith. Group Selection, *The Quarterly Review of Biology* 51: 277-283, 1976.

[Maynard SMith, 2002] J. Maynard Smith. Commentary on Kerr and Godfrey-Smith, *Biology and Philosophy* 17: 523-527, 2002.

[Maynard Smith and Szathmary, 1995] J. Maynard Smith and E. Szathmary. *The Major Transitions of Life*. New York: Oxford University Press, 1995.

[Mayr and Provine, 1980] E. Mayr and W. B. Provine (editors). *The Evolutionary Synthesis: Perspectives on the Unification of Biology*. Cambridge, MA: Harvard University Press, 1980.

[Michod, 1999] R. Michod. *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton, NJ: Princeton University Press, 1999.

[Nesse and Williams, 1994] R. M. Nesse and G. C. Williams. *Why We Get Sick: The New Science of Darwinian Medicine*. New York: Times Books, 1994.
[Okasha, 2001] S. Okasha. Why Won't the Group Selection Controversy Go Away? *British Journal for the Philosophy of Science* 52: 25-50, 2001.
[Okasha, 2002] S. Okasha. Genetic Relatedness and the Evolution of Altruism, *Philosophy of Science* 69: 138-149. 2002.
[Okasha, 2007] S. Okasha. *Evolution and the Levels of Selection*. New York: Oxford University Press, 2007.
[Parer, 1995] I. Parer. Relationship Between Survival Rate and Survival Time of Rabbits, *Oryctolagus cuniculus* (L.), Challenged with Myxoma Virus, *Australian Journal of Zoology* 43: 303-311, 1995.
[Parer *et al.*, 1995] I. Parer, W.R. Sobey, D. Conolly, and R. Morton. Sire Transmission of Acquired Resistance to Myxomatosis, *Australian Journal of Zoology* 43: 459-465, 1995.
[Park, 1932] T. Park. Studies in Population Physiology: The Relation of Numbers to Initial Population Growth in the Flour Beetle *Tribolium Confusum* Duval, *Ecology* 13: 172-181, 1932.
[Reeve, 2000] H. K. Reeve. Multi-Level Selection and Human Cooperation, *Evolution and Human Behavior* 21: 65-72, 2000. [Review of Sober and Wilson 1998].
[Reeve and Keller, 1999] H. K. Reeve and L. Keller. Levels of Selection: Burying the Units-of-Selection Debate and Unearthing the Crucial New Issues. In L. Keller (editor), *Levels of Selection in Evolution*. Princeton, NJ: Princeton University Press, 1999.
[Schmid-Hempel, 1998] P. Schmid-Hempel. *Parasites in Social Insects*. Princeton, NJ: Princeton University Press, 1998.
[Schrödinger, 1944] E. Schrödinger. *What is Life?* Cambridge: Cambridge University Press, 1944.
[Seeley, 1995] D. Seeley. *The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies*. Cambridge, MA: Harvard UP, 1995.
[Sober, 1984] E. Sober. *The Nature of Selection*. Cambridge, MA: MIT Press, 1984.
[Sober, 1991] E. Sober. Models of Cultural Evolution. In P. Griffiths (editor), *Trees of Life*. Cambridge, MA: MIT Press, 1991.
[Sober, 1993] E. Sober. *Philosophy of Biology*. Boulder, CO: Westview Press, 1993.
[Sober and Lewontin, 1982] E. Sober and R. Lewontin. Artifact, Cause, and Genic Selection, *Philosophy of Science* 49: 157-180, 1982.
[Sober and Wilson, 1994] E. Sober and D.S. Wilson. A Critical Review of Philosophical Work on the Units of Selection Problem, *Philosophy of Science* 61: 534-555, 1994.
[Sober and Wilson, 1998] E. Sober and D.S. Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press, 1998.
[Sober and Wilson, 2002] E. Sober and D.S. Wilson. Perspectives and Parameterizations: Commentary on Benjamin Kerr and Peter Godfrey-Smith's 'Individualist and Multi-Level Perspectives on Selection in Structured Populations', *Biology and Philosophy* 17: 529-537, 2002.
[Stanley, 1979] S. M. Stanley. *Macroevolution: Pattern and Process*. San Francisco, CA: W.H. Freeman, 1979.
[Stanley, 1981] S. M. Stanley. *The New Evolutionary Timetable: Fossils, Genes, and the Origin of Species*. New York: Basic Books, 1981.
[Sterelny, 1996] K. Sterelny. The Return of the Group, *Philosophy of Science* 63: 562-584, 1996.
[Sterelny, 2003] K. Sterelny. Last Will and Testament: Stephen Jay Gould's *The Structure of Evolutionary Theory*, *Philosophy of Science* 70: 255-263, 2003.
[Sterelny and Griffiths, 1999] K. Sterelny and P.E. Griffiths. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press, 1999.
[Sterelny and Kitcher, 1988] K. Sterelny and P. Kitcher. The Return of the Gene, *Journal of Philosophy* 85: 339-361, 1988.
[Trivers, 1971] R. Trivers. The Evolution of Reciprocal Altruism, *Quarterly Review of Biology* 46: 35-57, 1971.
[Trivers, 1974] R. Trivers. Parent-Offspring Conflict, *American Zoologist* 14: 249-274, 1974.
[Vrba, 1986] E. S. Vrba. What is Species Selection? *Systematic Zoology* 33: 318-328, 1986.
[Vrba, 1989] E. S. Vrba. Levels of Selection and Sorting with Special Reference to the Species Level. In P.H. Harvey and L. Partridge (editors) , *Oxford Surveys in Evolutionary Biology, Volume 6*. Oxford: Oxford University Press, pp. 111-168, 1989.

[Vrba and Gould, 1986] E. S. Vrba and S.J. Gould. The Hierarchical Expansion of Sorting and Selection: Sorting and Selection Cannot Be Equated, *Paleobiology* 12: 217-228, 1986.
[Wade, 1977] M. J. Wade. An Experimental Study of Group Selection, *Evolution* 31: 134-153, 1977.
[Wade, 1978] M. J. Wade. A Critical Review of the Models of Group Selection, *Quarterly Review of Biology* 53: 101-114, 1978.
[Wade, 1980] M. J. Wade. An Experimental Study of Kin Selection, *Evolution* 34: 844-855, 1980.
[Wade, 1982] M. J. Wade. Group Selection: Migration and the Differentiation of Small Populations, *Evolution* 36: 944-961, 1982.
[Wade, 2001] M. J. Wade. Infectious Speciation, *Nature* 409: 675-677, 2001.
[Wheeler, 1911] W. M. Wheeler. The Ant-Colony as an Organism, 1911. Reprinted in his *Essays in Philosophical Biology*. Cambridge, MA: Harvard University Press, 1939.
[Wheeler, 1920] W. M. Wheeler. Termitadoxa, or Biology and Society, 1920. Reprinted in his *Essays in Philosophical Biology*. Cambridge, MA: Harvard University Press, 1939.
[Williams, 1991] C. K. Williams and R.J. Moore. Inheritance of Acquired Immunity to Myxomatosis, *Australian Journal of Zoology* 39: 307-311, 1991.
[Williams, 1966] G. C. Williams. *Adaptation and Natural Selection: A Critique of some Current Evolutionary Thought*. Princeton, NJ: Princeton, NJ, 1966.
[Williams, 1985] G. C. Williams. A Defense of Reductionism in Evolutionary Biology, in R. Dawkins and M. Ridley (editors), *Oxford Surveys in Evolutionary Biology*, vol. 2. Oxford: Oxford University Press, 1985.
[Williams, 1992] G. C. Williams. *Natural Selection: Domains, Levels and Challenges*. New York: Oxford University Press, 1992.
[Wilson, 1975] D. S. Wilson. A Theory of Group Selection, *Proceedings of the National Academy of Sciences USA* 72: 143-146, 1975.
[Wilson, 1977] D. S. Wilson. Structured Demes and the Evolution of Group Advantageous Traits, *American Naturalist* 111: 157-185, 1977.
[Wilson, 1980] D. S. Wilson. *The Natural Selection of Populations and Communities*. Menlo Park, CA: Benjamin Cummins, 1980.
[Wilson, 1983] D. S. Wilson. The Group Selection Controversy: History and Current Status, *Annual Review of Ecology and Systematics* 14: 159-187, 1983.
[Wilson, 1988] D. S. Wilson. Holism and Reductionism in Evolutionary Ecology, *Oikos* 53: 269-273, 1988.
[Wilson, 1989] D. S. Wilson. Levels of Selection: An Alternative to Individualism in the Human Sciences, *Social Networks* 11: 257-272. Reprinted in E. Sober (editor), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 1989. 2^{nd} edition, 1993, pp. 143-154.
[Wilson, 1997a] D. S. Wilson. Altruism and Organism: Disentangling the Themes of Multilevel Selection Theory, *American Naturalist* 150 (supp.): S122-S134, 1997.
[Wilson, 1997b] D. S. Wilson. Incorporating Group Selection into the Adaptationist Program: A Case Study Involving Human Decision Making, in J. Simpson and D. Kendrick (editors), *Evolutionary Social Psychology*. Hillsdale, NJ: Erlbaum, 1997.
[Wilson, 1999] D. S. Wilson. A Critique of R.D. Alexander's Views on Group Selection, *Biology and Philosophy* 14: 431-449, 1999.
[Wilson, 2002] D. S. Wilson. *Darwin's Cathedral: Evolution, Religion, and the Nature of Society*. Chicago: University of Chicago Press, 2002.
[Wilson and Dugatkin, 1992] D. S. Wilson and L. Dugatkin. Altruism: Contemporary Debates, E. F. Keller and E.A. Lloyd (editors), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, 1992.
[Wilson and Sober, 1989] D. S. Wilson and E. Sober. Reviving the Superorganism, *Journal of Theoretical Biology* 136: 337-356, 1989.
[Wilson and Sober, 1994] D. S. Wilson and E. Sober. Reintroducing Group Selection to the Human Behavioral Sciences, *Behavioral and Brain Sciences* 17: 585-654, 1994.
[Wilson, 1971] E. O. Wilson. *The Insect Societies*. Cambridge, MA: Belknap Press, 1971.
[Wilson, 2000] E. O. Wilson. *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press, 2000; 2nd edition. 1^{st} edition, 1975.
[Wilson, 2003] R. A. Wilson. Pluralism, Entwinement, and the Levels of Selection, *Philosophy of Science* 70: 531-552, 2003.

[Wilson, 2004] R. A. Wilson. Test Cases, Resolvability, and Group Selection: A Critical Examination of the Myxoma Case, *Philosophy of Science* 71: 380-401, 2004.
[Wilson, 2005] R. A. Wilson. *Genes and the Agents of Life: The Individual in the Fragile Sciences: Biology*. New York: Cambridge University Press, 2005.
[Wolf *et al.*, 2004] J. B. Wolf, E.D. Brodie III, and M.J. Wade. Evolution When the Environment Contains Genes. In T. DeWitt and S. M. Scheiner (editors), *Phenotypic Plasticity: Functional and Conceptual Approaches*. New York: Oxford University Press, 2004.
[Wright, 1929] S. Wright. Evolution in a Mendelian Population, *Anatomical Record* 44: 287, 1929.
[Wright, 1931] S. Wright. Evolution in Mendelian Populations, *Genetics* 16: 97-159, 1931.
[Wright, 1968] S. Wright. *Evolution and the Genetics of Populations, vol. 1: Genetic and Biometric Foundations*. Chicago: University of Chicago Press, 1968.
[Wright, 1969] S. Wright. *Evolution and the Genetics of Populations, vol. 2: The Theory of Gene Frequencies*. Chicago: University of Chicago Press, 1969.
[Wright, 1977] S. Wright. *Evolution and the Genetics of Populations, vol. 3: Experimental Results and Evolutionary Deductions*. Chicago: University of Chicago Press, 1977.
[Wright, 1978] S. Wright. *Evolution and the Genetics of Populations, vol. 4: Variability Within and Among Natural Populations*. Chicago: University of Chicago Press, 1978.
[Wright, 1980] S. Wright. Genic and Organismic Selection, *Evolution* 34: 825-843, 1980.
[Wynne-Edwards, 1962] V. C. Wynne-Edwards. *Animal Dispersion in Relation to Social Behaviour*. Edinburgh: Oliver and Boyd.
[Wynne-Edwards, 1963] W. C. Wynne-Edwards. Intergroup Selection in the Evolution of Social Systems, *Nature* 200: 623-626, 1963.

WHAT IS EVOLVABILITY?

Kim Sterelny

1 THE METAZOA AND THE VOLVOCACEANS: TWO CONTRASTING FATES

For free-living unicellular protists, cell division is reproduction. A cell divides, and the two daughters go their separate ways. But very occasionally, something else happens. Somewhere between 700 and 800 million years ago, in a protist lineage closely related to the living choanoflagellates, cell division resulted in cell aggregation [King, 2004]. Instead of that division resulting in two daughter cells drifting off to their separate fates, the aggregation stayed together and shared a common fate. That lineage prospered. We can only conjecture at the reasons for its success: one possibility is simply that it was too large to engulf; its size made it safe from other protists. But though we do not know the advantage conferred by this experiment in collective life, we know its consequences. For I have just described the origins of the Metazoa: the lineage of multi-celled animals. So this aggregation was ancestor to a hugely diverse and hugely disparate clade.

The metazoans are one of the three great radiations of multi-celled life: the others are the green plants and the fungi. But these are by no means the only multi-celled lineages. There are different estimates in the literature, but no-one doubts that multicellularity has evolved repeatedly; more than a dozen times [Bonner, 1998; King, 2004]. So consider the contrast between these highly disparate lineages and another and more recent experiment. The plants evolved from green algae [Graham *et al.*, 2000]. But they are not the only multi-celled descendants of green algae. About 75 million years ago another experiment in collective life began: that of the volvocaceans. The founder of this lineage lived in shallow ephemeral ponds (if the ecology of its descendants is any guide). Phosphate is the crucial limiting resource in these ponds, and there is reason to suppose that multicellular volvocaceans store phosphorous more successfully that their single-celled competitors because they can store it in the extracellular matrix that fills the spaces between their component cells.

There has been a modest radiation of multicellular descendants of these protists [Kirk, 2001; Kirk, 2003]. Some multicelled lineages are halfway houses, intermediate between co-operative groups of individual cells, cells which retain independent reproductive fates, and structured collectives. There are volvocaceans like this, where cells stay together for a period before fissuring so that each component cell has its own separate shot at reproduction (as in *Chlamydomonas*). But other

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

volvocaceans are clearly structured collectives. In *Volvox carteri* there is a clear specialisation between somatic cells that enable individual volvocaceans to move through the water and stay at the surface (where they photosynthesise) and a much smaller number of much larger reproductive cells. True multi-celled organisms, not just co-operation between individual protists, have evolved from this protist lineage. Yet we find nothing like the complexity, the disparity, or the diversity of the metazoans. *Volvox carteri* is the upper limit of volvocacean complexity.

What explains this contrast between the metazoans and the volvocaceans? Questions of this kind can be asked at any grain. I have just contrasted the fate of two lineages that have experimented with multicellularity, but more specific versions of such questions arise within these lineages. Why, for example, is sex so developmentally plastic in many fish lineages but never in terrestrial tetrapods? Terrestrial vertebrates, unlike fish, never change sex in response to social cues. Lineages seem to differ from one another not just in their specific trajectories but also in the space of evolutionary possibility to which they have access. It just does not seem open to mammals to evolve the capacity to change sex in response to (say) skewed sex ratios, however advantageous that would be. Lineages contrast in the *actual pattern* of their evolutionary histories: some are strikingly more disparate (overall, or in particular respects) than other apparently comparable lineages. In some cases, it is plausible to suppose that these differences reflect differences in *evolutionary potential* rather than reflecting chance or selective environment.[1] There is a further temptation to think that these differences in evolutionary potential themselves have an explanation in the developmental biology of the organisms concerned: lineages differ at a time and over time in *evolvability*.

2 LIMITS ON VOLVOCACEAN DISPARITY

As noted above, multi-cellularity has evolved repeatedly, but only in a few lineages has this lead to a major radiation of multi-celled forms. My project in this paper is ask whether the idea of evolvability helps in explaining the different clade geometries of rich multi-celled lineages and relatively depauperate ones like the volvocaceans.[2] The volvocaceans are a good probe for investigating this problem, because we have a concrete hypothesis which explains the limits of volvox disparity. So let me begin by outlining that hypothesis. We can then ask: what do considerations of evolvability add to it?

The crucial idea is that while *Volvox carteri* has an established soma/germline distinction, it has established that distinction by a mechanism that limits somatic growth potential and (perhaps) the potential evolution of new cell types. In *Volvox*, the founding cell grows before cell division. The series of cell divisions

[1] Thus no marine vertebrate has hands: but in contrast to sexual plasticity, this contrast with terrestrial life seems to have a selective explanation.

[2] The mundane answer, of course, is time: we should remember that animal disparity may well have been unimpressive after the first seventy five million years of metazoan evolution.

that form the adult organism then largely take place without further growth. The initial process of cell division produces an embryo of 32 symmetrical cells. But at the next division half of the cells divide asymmetrically, producing a population of both large and small cells. The large cells from this process became the *gonadia* cells: the germ-line. The somatic ancestor cells, the smaller cells from this first asymmetric division, divide many more times and so *V. carteri* ends up with a couple of thousand small somatic cells and a handful of much larger gonadia cells. The whole show then turns inside out so that the small, somatic, flagellated cells are on the outside, and the larger immobile germ-line cells are on the inside, set in an extracellular matrix. The gonadia, alone, then grow so that each ends up roughly 1000 times the size of somatic cells. (See [Kirk, 2001] for a good discussion of these life history issues)

Single-celled green algae, the single-celled ancestor of the volvocaceans, live a two-phase existence: they have a motile growth phase followed by an immobile reproductive phase. Kirk suggests that the genes that controlled these phenotypic changes over time have been co-opted to build a soma/germ-line division at a time. The *lag*-gene when active, suppresses the development of flagella, eyespots and chemotaxis. It originally evolved to shift the single-celled ancestor from the growth phase to the reproductive phase. For no protist can divide in a flagellated form. The cellular machinery needed for flagella is also needed to control the even division of genetic material in mitosis [King, 2004]. So within gonadia cells, this gene is on. Within somatic cells, *lag* is off and *regA* is on. *RegA* is a mutation which disables chloroplast development. Cell division in *Volvox* depends on a size threshold: once somatic cells fall below this threshold, if their chloroplasts are disabled they cannot regrow to reach it again. The germ-line/soma division in *Volvox* thus depends on genetic mechanisms that ensure that somatic cells are small, and without the potential to grow. This limits division depth and hence organism size. It also probably limits the array of somatic cell types potentially available to this lineage.

Thus *Volvox* disparity is limited because somatic cells have a finite replication potential. The germ-soma division is organised through mechanisms which produce a somatic lineage which is restricted in total size, because the number of divisions depends on the size of founding gonadia, not on the resources harvested once division begins. Moreover, the policing mechanism which prevents somatic cells reverting to germ-line forms may restrict the potential range of cell morphology of somatic cells. The morphological architectures available in this lineage are thus sharply constrained by these restrictions on cellular constituents. Evolutionary potential is limited because the *Volvox* genotype-phenotype map links together three features of the *Volvox* developmental program: (i) establishing a soma-germ-line distinction; (ii) intrinsic limits on the generational depth of somatic cell lineages; (iii) limits on somatic cell diversity [Nedelcu and Michod, 2004]. There is no natural way to break the linkages between these features of volvocacean developmental programs.[3] One mechanism that has an important

[3]Especially, perhaps, because Volvox are haploid in their multi-celled life phase and hence are

role in policing outlaws — the establishment of the germ-soma distinction — also constrains somatic cell diversity.

3 FITNESS: A MODEL FOR EVOLVABILITY?

How does this picture of the regulation of cell differentiation relate to evolvability and to the idea that evolvability explains disparity and its limits? For surely we have just explained those limits without mentioning evolvability. There are two ways we might respond to this challenge, and the availability of these two responses exposes an ambiguity in the literature on evolvability. One idea is that the gene regulation explanation of *Volvox* disparity is shallow. The crucial issue is not whether there are *currently* mechanisms that constraint the range of cell architectures available in that lineage, but whether those mechanisms are themselves entrenched. Evolution depends on variation, and hence on the mechanisms which generate variation. Questions of evolvability are questions about the selectable phenotypic variation those mechanisms generate, and especially about how stable those mechanisms are over evolutionary change. To what extent do they themselves evolve, increasing or decreasing the variation that is available to selection?

Recent literature on bacterial evolvability has conceived of evolvability as a character that can itself evolve. In this literature, the focus has been on the evolution of mutation rates; in particular, whether elevated mutation rates show that there has been selection for increased mutation itself, or merely declining investment in error correction in impoverished environments.[4] Selection seems to favour elevated rates of mutation (and/or recombination) in stressful conditions. But that might just show that in tough times the price of more accurate replication is unaffordable. An increased mutation rate would then be an unfortunate side-effect of a stress-resistant thrifty phenotype. The case for thinking that selection might actually favour increased mutation rate is strengthened by the discovery of facultative mechanisms that increase mutation rates only in stressful conditions and then return to higher fidelity replication. These are inducible mutators: mutations which disrupt error-correction mechanisms. The case for treating these mutators as adaptations is stronger still if elevated mutation rates are targeted to specific regions of bacterial genomes. Radman, Matic and Taddei claim there are such systems. When microsatellite runs (repeated sequences of the same base pairs) are included within gene, the rate of initial copying errors increases. Normally, these errors are unimportant, for they are reliably corrected when repair mechanisms are working with their normal fidelity. But when these mechanisms are turned off, these microsatellite infested genes, and only these genes, become hypermutable [Radman *et al.*, 1999]. By restricting increased variation to these genes, the cost of an elevated mutation rate is reduced.

vulnerable to genetic noise: mutations are not masked by working copies of the mutated allele.

[4]See [Radman *et al.*, 1999] and [Earl and Deem, 2004]; for an overview, see [Chicurel, 2001].

All other things being equal, increases in the supply of variation increase the evolutionary response of a lineage to selection. Mutation rate is a component of variability. But even if selection favours elevated mutation rates, and mutation rates correlate with evolvability, this would not show evolvability to be a selectable character. My main reason for this claim is the idea that evolvability depends critically on population structure and environment. It is a characteristic of lineages, not organisms, for population structure concerns the division of a species into groups, and the flow of genes between those groups. I shall defend the idea that population structure is crucial to evolvability shortly. However, it is also true that the bacterial evolvability literature does not clearly distinguish *evolvability* from *phenotypic plasticity*. Evolvability is a property of a lineage. Plasticity is variously thought of as a property of an individual organism or a genome. A plastic genome maps onto different phenotypes in different environments. A plastic organism develops a different phenotype in different environments. For us, but not for bacteria, these two pictures of plasticity are equivalent. For in contrast to lumbering macrobes like ourselves, bacterial genomes and mechanisms of gene expression are not fixed over the life of individual organisms: lateral gene transfer is common in bacteria and decoupled from reproduction. Gene-changing might be a mechanisms of phenotypic plasticity, not of offspring variability.

If we are inclined to accept Janzen's argument that clones should be thought of as a single evolutionary individual, the distinction between evolvability and phenotypic plasticity becomes especially problematic [Janzen, 1977]. Dawkins argued convincingly against this view in the final chapter of [Dawkins, 1982]. But that argument applies only to multi-celled clonelines, for it depends on the importance of a single-celled bottleneck through which reproduction is channelled. The existence of that bottleneck allowed Dawkins to draw an important distinction between somatic and germ-line mutations, even within clonally reproducing organisms. No such distinction can be drawn for bacteria, and hence the status of a cloneline is unresolved. But even if we think of the single cell as the individual organism, the literature on bacterial evolvability conflates it with plasticity. For example, Partridge and Barton argue that a yeast prion is a replicator that enhances yeast evolvability. It does so by changing the way messenger RNA is read. The normal form of this protein binds to the stop codon, causing translation to cease. But the variant, prion-form often fails to bind. It thus results in the translation of messenger RNA into longer protein sequences. This normally depresses the fitness of yeast with the prion. But when conditions are bad, yeast with the variant, prion-form protein sometimes do better, thus increasing the frequency of this form of protein in the yeast population. There is selection for the prion-carrying yeasts in uncertain and fluctuating environments [Partridge and Barton, 2000]. But though Partridge and Barton call this selection for evolvability, it is actually selection for phenotypic plasticity. Prion-carrying yeast alter their own phenotypes, not just the downstream phenotypes of their daughter cells. Yeast that cannot anticipate their own environment are individually better off, on average, if they carry the prion-protein. While most of them will do marginally worse than the wild-type, a

few will do much better.

In my view, evolvability is not a selectable trait, even though mutation rate may be. There is an alternative and better way of thinking of evolvability. Think of it as akin to fitness. Fitness is not a trait of an organism; the fitness of a centipede is not like its segmentation pattern or leg number. This view of fitness is not completely uncontroversial. For fitness has been understood as an explanatory property: as a measure of the congruence between an organism and its circumstances. The idea is that variations in congruence explain variations in reproductive success, both within and across populations. But it has proved difficult to specify a general congruence relation, for it would have to abstract away from the specific details of both an organism and its world. Moreover the fitness of an organism is sensitive not just to the environment but to population structure. Fitness is not just a relationship between organism and environment. The literature on the evolution of co-operation has made this clear: the fitness that is relevant to evolutionary dynamics is relative fitness, not absolute fitness. Defecting traits which lower the absolute fitness of every agent in the population can invade, so long as their effect on those without the trait is more severe than the effect on those with it. Moreover the fitness implications of particular patterns of social behaviour often depend both on the role of the agent within the group of which it is a part, and on the role of the group within the population. Population structure helps explain relative fitness [Kerr and Godfrey-Smith, 2002]. Thus fitness is now typically understood as a dispositional property of organisms. Fitter members of a population are disposed to have more reproductive success than their rivals. But if we want to explain this variance, we appeal to the specific features of the organisms' phenotypes. It is these that explain success or failure.

However, even if fitness is not a character state that explains success or failure, it does not follow that there is nothing interesting to say in the language of fitness. There is an intermediate level of generality between the ascription of fitness differences to organisms and the analysis of specific phenotype/environment couplings. Thus within evolutionary biology we distinguish between the contribution of natural and sexual selection to organism success. In understanding the evolution of co-operative behaviour, it is crucial to distinguish between fitness effects that derive from within-group differences and those that derive from cross-group differences. We distinguish between frequency-dependent aspects of fitness and fitness deriving from optimisation. If a finch's beak is optimised to the specific seasonal conditions it encounters, the benefit it derives is insensitive to others' phenotypes. Contrast this with the female-mimicking mating strategy in the giant cuttlefish. Small males often assume the body shape and patterns of females at breeding aggregations, lurking near a breeding pair, and relying on the larger guarding male being distracted by another intruding male. When there is such a distraction, the female mimic resumes a male appearance and attempts to mate, often successfully. In contrast to the finch, this strategy depends on the supply of intruding males. These intruders disrupt male guarding and allow female mimics their opportunity to mate [Norman *et al.*, 1999]. Sexual selection, population-

structured selection, and frequency-dependent selection are conceptual tools that mark important aspects of the explanation of success and failure. Yet they abstract away from specific features of organisms and their environments.

4 EVOLVABILITY, INDIVIDUALS AND ENVIRONMENTS

I suggest that fitness understood this way is a model for understanding evolvability. Just as fitness is a dispositional feature of individual organisms, evolvability is a dispositional feature of lineages. It is not a character of an organism or its developmental system; it is not a trait. It is true that the evolvability of a lineage depends on the developmental programs of the individual organisms in the lineage. Moreover most of the literature on evolvability has focused on internal factors — features of individual developmental systems — and their role in either generating or constraining variation on which selection acts.[5] This is certainly true of Kirschner's and Gerhart's review article that returned evolvability to the agenda of evolutionary biology [Kirschner and Gerhart, 1998]. This review was about variation, and in particular mechanisms that reduced the cost of variation by making them less likely to be catastrophic. For example, Kirschner and Gerhart bring out the crucial connection between evolvability and phenotypic plasticity. Without plasticity, evolutionary change would require impossibly co-ordinated genetic changes. Think, for example, of the evolution of new cell morphologies. Mitosis is impossible unless microtubes connect to the chromosomes in the cell and mediate their symmetrical segregation to the spindle poles at each end of the cell. Microtubes find chromosomes by mechanisms of undirected variation and selective retention: the tube is unstable and is reabsorbed unless it happens to connect with a chromosome, in which case it stabilises. Through this mechanism, the process of chromosome sorting does not have be pre-programmed with information about the number of chromosomes, their location, or of the size and shape of the cell. This mechanism does not need to be genetically re-tuned if other mutations change the cell architecture [Kirschner and Gerhart, 1998, 8422]. These ideas are important: features of individual developmental systems are crucial to lineage-level evolvability. Indeed, in my earlier work on evolvability I attempted to specify the general characteristics of inheritance and development that promote lineage-level evolvability [Sterelny, 2001; 2004].

However, it is crucial to see that the evolvability of a lineage depends on much more than the developmental programs of individual organisms within the population, and the extent to which those programs vary.

1. The environment matters: for variation in the environment determines the fraction of the reaction norm that is expressed and thus exposed to selection.

[5]The exception is the literature on non-genetic inheritance and its role in evolution. That literature has been much more sensitive to the role of the environment in the reliable transmission of parent-offspring similarities [Avital and Jablonka, 2000; [Odling-Smee *et al.*, 2003].

More uniform environments allow cryptic genetic variation to survive unexpressed; variation that may become important if the environment changes. This point has been developed in an important recent paper by Suzanna Rutherford. Populations have unexpressed genetic variability: for example, in natural *Drosophila melanogaster* populations, she argues, there are on average hundreds of thousands of base pair differences between the average haploid genotype. Yet these are strikingly phenotypically uniform populations. Much genetic variation is effectively neutral because it does not give rise to phenotypic variation. Yet though this variation is cryptic while the environment is stable, it can be unmasked. It is *unexpressed* difference, not *inexpressible* difference. Consider, for example, the two forms of the human *CYP1A1* gene. In non-smokers, these two forms are phenotypically equivalent. But they are associated with marked difference in lung cancer risk for smokers. In particular, one form of the gene makes moderate smoking much more dangerous than it would otherwise be. Likewise, new genetic variation can cause previously silent differences to be expressed. A mutant form of the heat-shock protein *Hsp90* in *Drosophila* unmasks mutations in other genes which would otherwise be silent [Rutherford, 2000].

Factors which make the environment to which a lineage is exposed more uniform are therefore important, for they allow genetic variation to be stored. In masking existing variation, they limit microevolutionary response to local variation but thereby enhance long-run evolvability by preserving genetic variation which would otherwise be eliminated from the gene pool. Moreover, there are such masking mechanisms: for organisms often have just these effects on their own environment. Often organisms in part engineer the developmental environment of the next generation, and thus make them more homogenous than they would otherwise be. Mistletoe seeds germinate only after passing through the digestive system of the mistletoe bird; the seeds of many Australian plants germinate only after fires to which their parents contribute [Odling-Smee *et al.*, 2003].

2. Population structure is relevant to the distribution of the genetic resources of the species. This is most vividly illustrated by prokaryote populations. For though prokaryotes have limited chromosomal evolution (their chromosome is circular, so there is no recombination), there is rich horizontal transfer of ready-made genetic material. Plasmids, phage DNA and transposons are all mechanisms of horizontal gene movement, with different size packets. Given the ubiquity of horizontal gene transfer, the richness of local genetic resources is obviously important [Carroll, 2002]. But though this is a very vivid case where evolutionary response depends on resources available in the local population, the general moral applies to the more familiar world of the macrobes. Microevolutionary change takes place within local populations, and if these are isolated from one another, there may well be potentially important gene combinations which are unavailable, because the variants

that would form the combination have arisen in different populations.

3. Population structure is relevant to the type of microevolutionary changes that are likely. It is a truism of evolutionary theory that microevolutionary change in a species divided into many small, isolated populations is likely to be quite different from microevolutionary change in a species divided into a few large ones. Drift and founder effects are likely to be important in the first case but not the second. Population structure is also crucial to multi-level selection. For selection at the level of groups can be important only if a species is divided into groups that vary one from another in ways important to their productivity. Multi-level selection probably plays a crucial role in the evolution of new levels of biological organization, and hence is of critical importance in expanding the space of evolutionary possibility [Michod, 1999; Kerr and Godfrey-Smith, 2002].

4. Population structure is relevant to the extent to which microevolutionary change is buffered or protected. As Eldredge has repeatedly emphasised, local adaptation in local populations is fragile; it is easily lost if demic structure is unstable, for migration from other populations breaks up gene combinations co-adapted to local circumstances. A population can stably adapt to its local world only if it extrinsically or intrinsically isolated [Eldredge, 2003].

The crucial point to remember is that significant evolutionary change typically depends on the accumulation of smaller changes. Accumulation, in turn, depends on the existence of ratchets that protect each small advance. There is an internal ratchet: high fidelity replication preserves rare favourable variations. But while internal, organism-level ratchets are necessary for evolvability, they are not sufficient. Potentially favourable change must be *preserved* by individual level inheritance processes. But these must then be *amplified* in the population and then protected (from being swamped by immigration) by population level processes, thus making a further mutational change and a further iteration far more likely. Selection is not just a passive consumer of variation: it is a creator of variation. When a phenotypic change from P to P* depends on a sequence of genetic changes from $G_0 \Rightarrow G_1 \Rightarrow G_2$, selection can make G_2, vastly more probable by amplifying the frequency of G_1. This makes the features of environment and population structure that are relevant to cumulative selection equally relevant to evolvability. Evolvability is not a characteristic of individuals: it emerges from an interplay between (i) individuals and their developmental systems, (ii) the populations and lineages of which they are a part, and (iii) the environments within which they are embedded.

Components of Evolvability If evolvability is a dispositional feature of a lineage, there is an important sense in which differences in evolvability do not explain differences in lineage disparity. To say that the basal metazoan lineage was highly evolvable does not explain the evolutionary radiation of the animals; likewise, to say that the volvocaceans have very limited evolvability explains nothing about

its limited disparity. But just as there is middle-level theory of the sources and consequences of fitness differences, there may be a middle-level theory of evolvability: characteristics shared by many lineages, and which play similar roles in explaining dispositions to evolve over time. The right way to understand proposals about evolvability is to see them as proposals about such components of evolvability: proposals about the evolution or effects of such components. I will illustrate this idea by discussing two major foci of the evolvability literature. One is the explanation of continuing evolutionary plasticity over time. William Wimsatt has pointed out that early aspects of development, in particular, should be increasingly hard to alter. For they are causally relevant to many downstream developmental stages; and the more downstream consequences a particular mechanism has, the more likely it is that changing that mechanism will have some disastrous consequence [Wimsatt and Schank, 1988]. So we should expect ontogenies to be developmentally entrenched, except perhaps for their terminal stages. And yet they are not [Raff, 1996]. The second is the mirror-image of continued flexibility: the existence and explanation of phenotypic gaps. In discussing these familiar ideas, I hope to illustrate two basic points. First: just as with fitness, we have developed conceptual tools that mark important aspects of evolutionary flexibility and inflexibility; tools that abstract away from specific features of organisms and their environments. Second: there is a serious imbalance in this work. For despite the importance to evolvability of environment and population structure, this discussion is almost entirely focused on internal aspects of evolvability.

Let me begin with a much-discussed example: modularity. A developmental module is a system that develops relatively independently of other such systems. A group of genes make up a genetic module, if there are many interactions between them and their products, and relatively few interactions between these genes and others. Developmental modules are also evolutionary modules [Brandon, 1999]: if a trait (say, tooth shape) *develops* independently of other traits, variation in that trait will be independent of variation in other traits. And if that is true, tooth shape can respond to selection without thereby altering other aspects of the phenotype. If the evolution of tooth shape is decoupled from the evolution of other traits, tooth shape will be more responsive to change in selective regimes. For it will be able to change even when there is stabilising selection on other aspects of the phenotype. The idea there is an important connection between modularity and evolvability dates to a classic paper of Lewontin, in which he argued that lineages could respond to selection only to the extent that their characters were "quasi-independent". Tooth shape and tooth number are quasi-independent only if each can vary independently of the other. If instead, their evolutionary fate is coupled, it will be impossible to select for (say) high-crowned teeth without changes in tooth number [Lewontin, 1978]. In such circumstances, it is not sufficient for high crowns to be selectively advantageous. Such teeth will evolve only if this change is worth the price in fitness of reduced tooth number.

The point that evolutionary response to environmental change is sensitive to linkage between traits has not been controversial, though it is difficult to test em-

pirically. For the extent to which traits develop independently of one another is not an obvious feature of organism phenotypes. Comparison between sister groups is one way to show that traits do, or do not, vary independently of one another. Finlay and her colleagues have argued that mammalian brain structures do not show much modularity: the olfactory bulb seems to be able to shrink or grow independently of size change in other regions, but this is the exception rather than the rule [Finlay *et al.*, 2001]. In a more explicit test of the link between modularity and evolvability, Yang compared two insect sister groups: the hemimetabolous insects and the holometabolous insects. Holometabolous insects undergo full metamorphosis, and hence the morphology of the larvae is decoupled from the morphology of the adult into which the larvae eventually develops, and hence are often both morphologically and ecologically quite different from those adult forms. In contrast, the nymphs of hemimetabolous insects are quite similar to the adults they will become. It is then perhaps no coincidence that the holometabolous insects are vastly the more species rich of the two clades. This difference in diversity makes sense in the light of the developmental difference. In the more diverse clade, the morphology of larvae is decoupled from that of the adult, and this allows the larvae to differentiate ecologically and morphologically from the adults: they can thus avoid competing with their own adult forms, and their adaptation to their own life ways is not constrained by adult adaptation to adult lifeways.

So there has been some attempt to test the link between modularity and evolvability. But there has been more focus on the extent to which developmental programs are modular, and the ways in which modularity changes over time. Gunter Wagner and his co-workers are responsible for developing models of modularity. In these models, selection will reduce epistatic linkages between two sets of genes and their associated traits when there is directional selection on one trait and stabilising selection on the other. Such a regime will select for modifier genes which suppress epistatic connections between the stabilised trait and the evolving trait. Thus we get selection for modularity whenever genes have pleiotropic effects with opposite fitness values. This may well be quite common [Wagner and Altenberg, 1996; Wagner *et al.*, 1997]. Somewhat less obviously Wagner has pointed out that modularity can increase as a side-effect of selection for genetic canalisation. The more genetic inputs there are to trait T, the more opportunities there are for T to be perturbed by genetic and developmental noise. So the development of a trait can be canalised by making its development sensitive to fewer genetic inputs [Wagner *et al.*, 2005].

So modularity reduces developmental entrenchment: selection regimes of the kinds Wagner characterises preserves and even increases the extent to which traits are quasi-independent, and hence help explain how it is that phenotype space has been extensively explored. Thus Roger Thomas has defined a seven-dimensional skeleton space, a space that specifies the array of possible skeleton types. In his view, this space is explored richly and quickly, because the elements of skeleton-design are module-like in their structure: for example, the materials from which these skeletal elements can be built are either rigid or flexible, and this design

decision varies independently of another: whether the distinct elements are physically unconnected, jointed or fused. The result of modular design is rapid and rich exploration. As he counts them, the Metazoa, living and extinct, have employed about 180 skeleton designs; 146 of these are found in the Burgess shale metazoans [Thomas, 2005]. If we concentrate on examples like this, we will think that the explanatory agenda is that of explaining the richness of evolutionary possibilities open to lineages.

In contrast, the developmental constraints literature has an opposite focus: that of explaining surprising limits on disparity within a lineage. In a series of recent works, Wallace Arthur has developed a particularly clear and thoughtful version of this conception of the relevance of development to evolution, with some striking examples of odd gaps in nature: gaps that seem unlikely to have a selective explanation. But though I begin with gaps, his is not just a theory of why some regions of phenotype space are empty. Consider, for example, the missing centipedes [Arthur, 2000, 54]. Geophilomorphian centipedes are a family of a thousand or so species, and there is a lot of variation in the number of segments into which their trunks are divided: these centipedes have from 29 to 191 segments in their trunks. But segment patterns come only in odd numbers: so where are the geophilomorphians with, say, trunks of 66 segments? The Lithobiomorphian centipedes are even more restricted: there are a thousand or so species of such centipedes, and they all have trunks divided into fifteen segments. In this family, there is no extant variation at all. It is hard to imagine an environmental factor that would penalise variation from these patterns, so it is natural to suspect that limits in the supply of variation explain such holes in the fabric of disparity. In the light of such examples, Arthur thinks that developmental bias — features of development which make some changes from current phenotypes probable, others possible but improbable, and other impossible, are important in explaining the direction of evolution. The supply of variation plays an essential role in determining the evolutionary trajectories, and not just be constraining the space of phenotypic possibility in unexpected ways.

No-one denies, of course, that a supply of variation is essential to evolutionary change. But it does not follow that the supply of variation makes a difference to the direction of change. If variation is typically densely and isotropically clustered around current phenotypes, then change over time will depend only on the selective environment, not the supply of variation. For variation would be available for whatever change selection favoured. We would expect variation to be isotropic if the genotype-phenotype map was one-to-one, and in which small variation in the existing genes for a trait would result in small variations in that trait. On the Fisherian conception of the relationship of genotypes to phenotypes, phenotype traits depend on many small-effect genes. On that model it might be reasonable to take the supply of variation to be irrelevant to the direction of phenotype change. But within population genetics it is no longer a standard to suppose that evolutionary change depends only on small-effect genes (see, for example, [Leroi, 2000] and [Orr, 2000]). Moreover, we have good reason to suspect that

the Fisherian model is not a good general account of the genotype-phenotype relation. Continuously variable traits might show Fisherian patterns of variation, with standing variation in the population generating normal variation on each side of the current mean. Beak length in Darwin's finches may well be influenced by a large number of small-effect genes. Genotypes close to actual genotypes would produce beaks differing slightly in depth, length and breadth from actual beaks. If that is right, there is a dense cluster of possible beaks varying slightly from actual beaks, and so selection can find a path through beak space in any direction by small increments. But many traits are not continuously variable in this way: changes in segment number from fifteen to sixteen cannot be understood as an incremental path through phenotype space. The same is true of other discontinuous changes: for example, the shift from the ancestral bilaterian symmetry pattern to the 5-fold symmetry of the echinoderms.

In short, there is no *a priori* reason to suppose that the supply of variation is typically unbiased. Arthur argues that the supply of variation is indeed importantly structured. In part, this is because he thinks there are failures of quasi-independence. Consider mammal leg length. Most mammals are close to symmetry, with front and back legs of near-equal length. Mammals with small front legs have small back legs, and so on as we move from shrew and mouse to dog to deer to giraffe. The pattern has exceptions: for example, kangaroos and their relatives have much longer and more robust hind legs than front legs. But there are no exceptions that run the other way: powerful forelegs and relatively dwarfed rear legs (though the Tasmanian devil comes close). So the pattern of rough equivalence between front and back looks quite robust. Perhaps it is just the result of selection: after all, it is not easy to think of ways of making a living that select for long front legs and short back ones. Arthur's alternative hypothesis is that this pattern is it the result of difficult-to-alter features of mammalian development programs. Adaptive complexes like that of the kangaroo are difficult to reach because of structure in the supply of variation: the natural supply of variation will not provide much variance in length between the two limb pairs, and hence adaptive peaks with unequal lengths are quite often too distant from current variational pools for them to be available. Crucially, though, Arthur's argument is not just one which appeals to incomplete modularity to explain missing phenotypes. In his view, some phenotypes are larger genetic targets, and hence are more likely to appear, than others. Some phenotypes are multiply realisable — different genetic changes will give rise to the same phenotypic outcome. Other phenotypes have a single genetic profile; their development depends on a specific set of genes. Those phenotypic variants that are reachable by a number of tracks through gene space are more probable than those that depend on a specific sequence of gene changes. So bias in the supply of variation need not depend on biases in mutation itself. The many/many nature of the genotype ⇒ phenotype map induces its own biases in variation.

Suppose that there are indeed biases in the structure of variation: Arthur argues that those biases matter, because evolutionary change is path-dependent.

Populations often find themselves on flat fitness landscapes near to but not on the slopes of incompatible local optima. Such populations will not evolve under selection until new variation extends the standing variation in the population to the foothills of one of the local optima. Once that variant has established in the population, selection will drive the population to the top of that local fitness peak, and the other will no longer be accessible to that population. The direction of evolution will depend on the order in which variations pop up in the population. In essence, the trajectory of a lineage through phenotype space will depend party on the supply of variation if, first, the supply of variation is not isotropic around whatever the current phenotype happens to be; and, second, phenotypic evolution as path-dependent. Biases in the supply of variation make some variants available for selection sooner, and others later. Order matters: selective response to the first variant shifts the population[6] in an adaptive landscape. While there would have been selection in favour of the second variant, had it arrived at the same time as the first, it is selectively penalised once evolution has shifted the population.

This sketch of modularity and of structured variation is not meant to be definitive. There are significant and unresolved problems with these ideas. One is the striking neglect of environmental contributions to evolvability; even when such contributions might help explain a problem with which the theorist is struggling. Consider, for example, Mark Ridley's interesting and important analysis of the evolution of complexity. Ridley regards complexity as a challenge for evolvability: for it seems that a lineage must already be complex before it can evolve complexity. Building complex phenotypes is information hungry. If that information is in the organism's genome, it follows that complex organisms require long genomes. But without complex policing and error correcting mechanisms (whose construction themselves require long genomes) long genomes tend to decay. Ridley concludes that complex life is "difficult" to evolve[7] [Ridley, 2000]. An obvious, but unexplored option is to consider the role the environment might play in buffering similarity across the generations. A second problem is the reliance in the discussion of both modularity and structured variation on the concept of a genotype $\Rightarrow$ phenotype map. That notion is problematic, and particularly so in this context. The metaphor of a genotype $\Rightarrow$ phenotype map understates the importance of non-genetic developmental resources to development and the context sensitivity of the effects of genes on phenotypes [Oyama *et al.*, 2001]. Those problems are important, but the idea of such a map is even more problematic in the context of evolutionary transitions in individuality. In thinking of, for example, the volvocaceans, is the phenotype a phenotype of individual cells or of a collective? However, despite these problems, the discussion of modularity and of structured variation illustrates the key idea of this section. In discussing the components of evolvability, we can profitably idealise way from the specific developmental and

[6]Moreover, the spread of variation is now around the new phenotype which means variants which were once quite probable become much less likely

[7]Though given his argument, it seems that he ought to have concluded that it is impossible to evolve

environmental contexts of specific lineages. The constraints on volvocacean morphological disparity are exemplified in different forms in other lineages as well.[8]

BIBLIOGRAPHY

[Arthur, 2000] W. Arthur. The concept of developmental reprogramming and the quest for an inclusive theory of evolutionary mechanisms. *Evolution and Development*, 2(1): 49–57, 2000.

[Avital and Jablonka, 2000] E. Avital and E. Jablonka. *Animal Traditions: Behavioural Inheritance in Evolution*. Cambridge, Cambridge University Press, 2000.

[Bonner, 1998] J. T. Bonner. Origins of multicellularity. *Integrative Biology*, 1(1): 27–36, 1998.

[Brandon, 1999] R. N. Brandon. The units of section revisited: The modules of selection. *Biology and Philosophy*, 14: 167–180, 1999.

[Carroll, 2002] R. L. Carroll. Evolution of the capacity to evolve. *Journal of Evolutionary Biology*, 15: 911–921, 2002.

[Chicurel, 2001] M. Chicurel. Can organisms speed their own evolution?. *Science*, 292(8 June 2001): 1824–1827, 2001.

[Dawkins, 1982] R. Dawkins. *The Extended Phenotype*. Oxford, Oxford University Press, 1982.

[Earl and Deem, 2004] D. J. Earl and M. W. Deem. Evolvability is a selectable trait. *Proceedings of the National Academy of Science*, 101(32): 11531–11536, 2004.

[Eldredge, 2003] N. Eldredge. The sloshing bucket: How the physical realm controls evolution. In J. P. Crutchfield and P. Schuster (eds.), *Evolutionary Dynamics: Exploring The Interplay of Selection, Accident, Neutrality and Function*. Oxford, Oxford University Press, 2003.

[Finlay *et al.*, 2001] B. L. Finlay, R. B. Darlington, and N. Nicastro. Developmental structure in brain evolution. *Behavioral and Brain Sciences*, 24: 263–308, 2001.

[Graham *et al.*, 2000] L. Graham, M. Cook, *et al.*. The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proceedings of the National Academy of Science*, 97(9): 4535–4540, 2000.

[Janzen, 1977] D. H. Janzen. What are dandelions and aphids? *The American Naturalist*, 111: 586–589, 1977.

[Kerr and Godfrey-Smith, 2002] B. Kerr and P. Godfrey-Smith. Individualist and multi-level perspectives on selection in structured populations. *Biology and Philosophy*, 17(4): 459–476, 2002.

[King, 2004] N. King. The unicellular ancestry of animal development. *Developmental Cell*, 7: 313–325, 2004.

[Kirk, 2001] D. Kirk. Germ-soma differentiation in *Volvox*. *Developmental Biology*, 238: 213–223, 2001.

[Kirk, 2003] D. Kirk. Seeking the ultimate and proximate causes of *Volvox* multicellularity and cellular differentiation. *Integrative and Comparative Biology*, 43(2): 247–253, 2003.

[Kirschner and Gerhart, 1998] M. Kirschner and J. Gerhart. Evolvability. *Proceedings of the National Academy of Science, USA*, 95: 8420–8427, 1998.

[Leroi, 2000] A. M. Leroi. The scale independence of evolution. *Evolution & Development*, 2(2): 67–77, 2000.

[Lewontin, 1978] R. C. Lewontin. Adaptation. *Scientific American*, 239: 156–169, 1978.

[Michod, 1999] R. E. Michod. *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton, Princeton University Press, 1999.

[Nedelcu and Michod, 2004] A. Nedelcu and R. E. Michod. Evolvability, modularity and individuality during the transition to multicellularity in volvocacean green algae. In G. Schlosser and G. P. Wagner (eds.), *Modularity in Development and Evolution*. Chicago, University of Chicago Press, pages 466–489, 2004.

[Norman *et al.*, 1999] M. Norman, J. Finn, and T. Tregenza. Female impersonation as an alternative reproductive strategy in giant cuttlefish. *Proceedings of the Royal Society, London, Series B*, 266: 1347–1349, 1999.

[Odling-Smee *et al.*, 2003] J. Odling-Smee, K. Laland, *et al.*. *Niche Construction: The Neglected Process in Evolution*. Princeton, Princeton University Press, 2003.

[8]Thanks to Brett Calcott, Patrick Forber and Mohan Matthen for their comments on earlier versions of this chapter.

[Orr, 2000] H. A. Orr. Adaptation and the cost of complexity. *Evolution*, 54(1): 13–20, 2000.
[Oyama *et al.*, 2001] S. Oyama, P. E. Griffiths, and R. D. Gray, eds. *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge, Mass, MIT Press, 2001.
[Partridge and Barton, 2000] L. Partridge and N. Barton. Evolving evolvability. *Nature*, 407(28 September): 457–458, 2000.
[Radman *et al.*, 1999] M. Radman, I. Matic, *et al.*. Evolution of evolvability. *Annals of New York Academy of Sciences*, 870: 146–155, 1999.
[Raff, 1996] R. Raff. *The Shape of Life: Genes, Development and the Evolution of Animal Form*. Chicago, Chicago University Press, 1996.
[Ridley, 2000] M. Ridley. *Mendel's Demon: Gene Justice and the Complexity of Life*. London, Weidenfeld and Nicholson, 2000.
[Rutherford, 2000] S. Rutherford. From genotype to phenotype: Buffering mechanisms and the storage of genetic information. *Bioessays*, 22(12): 1095–1105, 2000.
[Sterelny, 2001] K. Sterelny. Niche construction, developmental systems and the extended replicator. In S. Oyama, P. E. Griffiths and R. D. Gray, (eds.), *Cycles of Contingency*. Cambridge, MIT Press, pages 333–350, 2001.
[Sterelny, 2004] K. Sterelny. Symbiosis, evolvability and modularity. In G. Schlosser and G. Wagner (eds.), *Modularity in Development and Evolution*. Chicago, University of Chicago Press, pages 490–516, 2004.
[Thomas, 2005] R. D. Thomas. Hierarchical integration of modular structures in the evolution of animal skeletons. In W. Callebaut and D. Rasskin-Gutman (eds.), *Modularity: Understanding The Development and Evolution of Natural Complex Systems*. Cambridge, MIT Press, pages 239–258, 2005.
[Wagner and Altenberg, 1996] G. P. Wagner and L. Altenberg. Complex adaptations and the evolution of evolvability. *Evolution*, 50: 967–976, 1996.
[Wagner *et al.*, 1997] G. P. Wagner, G. Booth, and B. C. Homayoun. A population genetic theory of canalization. *Evolution*, 51(2): 329–347, 1997.
[Wagner *et al.*, 2005] G. P. Wagner, J. Mezey, and R. Calabretta. Natural selection and the origin of modules.In W. Callebaut and D. Rasskin-Gutman (eds.), *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. Cambridge, MIT Press, pages 33–50, 2005.
[Wimsatt and Schank, 1988] W. C. Wimsatt and J. C. Schank. Two constraints on the evolution of complex adaptations and the means of their avoidance. In M. H. Nitecki (ed.), *Evolutionary Progress*. Chicago, University of Chicago Press, pages 231–275, 1988.

DEVELOPMENT: THREE GRADES OF ONTOGENETIC INVOLVEMENT

D. M. Walsh

1 INTRODUCTION

Viktor Hamburger [1980] famously claimed that developmental biology was left out of the modern synthesis from its inception in the early 20^{th} Century. He didn't mean that some anti-ontogeny conspiracy met in some smoke-filled room and decreed that developmental biology should be excluded. At the time, developmental biology was a discipline in good standing, and those who forged the synthesis were well aware of it. Developmental biology was 'left out' simply because it didn't seem to matter much one way or another; it seemed to pose no threat, nor to offer any great enhancement to the emerging orthodoxy. In Hamburger's words, evolutionary biologists came to treat development as a 'black box', a process that played some contributory role in evolution, but whose details had little bearing on the correctness or otherwise of the synthesis theory. So for much of the 20^{th} Century developmental biology languished as a relative outsider among biological disciplines, an area of only marginal interest to evolutionists. It is only recently, now that more of the details of ontogeny are understood, that biologists have had cause to rethink the place of development in evolutionary biology. Developmental biology is now one of the most rapidly growing disciplines in biology. It has witnessed enormous advances in the understanding of the mechanics — genetic, epigenetic and environmental — of development in the last twenty years. It is clear now that an understanding of the processes of development is of cardinal importance to the project of explaining the mechanisms of evolution. Yet in spite of this flourishing — or perhaps because of it — there is little consensus on just how this newfound knowledge should impact our conception of evolutionary theory.

I attempt here to outline the space of possible roles for ontogeny in evolutionary biology. I don't intend to adjudicate — although I do have my preferences. It's an empirical issue which of the alternatives is most plausible. But I hope at least to help shed some light on just what that empirical issue is.

One coherent possibility is that an understanding of the processes of ontogeny should occasion no revision to the version of the modern synthesis theory we have grown so comfortable with [Raff, 1996]. But there are others. I develop two other, increasingly central, possible roles for the process of development in evolution. These, by degrees, challenge the standard version of the modern synthesis. In

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

developing the space of possibilities, it will help if we understand the rationale behind the original marginalization of ontogeny. The humble status of development more or less follows from the precepts of the modern synthesis, at least the version that was forged in the early 20^{th} Century, and which solidified in the latter half [Gould, 1983]. The more we understand about these precepts, on the one hand, and the mechanisms of development, on the other, the more apparent it should become that the marginalization of development from the theory of evolution is no longer tenable, if it ever was.

2 THE FRAGMENTATION OF EVOLUTION

Evolution is a process. Since the advent of the modern synthesis theory of the early 20^{th} Century it has come to be seen as an amalgam of three more or less distinct processes: (i) *Development*: an intra-organismal process by which genotype becomes phenotype, or by which an undifferentiated organism precursor becomes a highly complex, highly differentiated organism; (ii) *Inheritance*: an inter-organismal process that secures the resemblance of offspring to parent; (ii) *Adaptation*: a supra-organismal process, in which populations of organisms come to comprise individuals well suited to their conditions of existence. When there is variation with respect to the inherited traits of organisms, adaptive evolution will occur over time. Crucially for the modern synthesis, these processes are thought to have a significant degree of independence. The process of development within a generation doesn't affect the content of those traits transmitted between generations by the process of inheritance. The mechanisms by which traits are transmitted between generations, and develop within individuals, are indifferent to the adaptiveness of those traits. That is to say transmission and development are not what cause an individual's adaptedness to the conditions of its existence. At the same time, the process that promotes adaptation does not alter the content of the traits that are inherited and developed: a single iteration of selection doesn't change traits, it simply selects between them.

This fragmentation of processes is not a prerequisite for the occurrence of evolution; it is a theoretical commitment of the modern synthesis. It is possible to have a theory of evolution that does not dissociate these three processes so comprehensively. For example, in Lamarck's theory, inheritance, development and adaptation are all consequences of the same process. According to Lamarck, organisms develop in response to two distinct sorts of factors, an inner striving for the realization of their type and external environmental influences. Development adapts an organism to environmental pressures and the adaptive novelties thus generated are inherited. Inheritance, development and adaptation are intimately intertwined, consequences of the principles that guide ontogeny. Similarly, Haeckel's later biology, while formally consistent with Darwin's, is essentially a theory of the evolution heritable ontogenies. Diversity is explained by the Biogenic law; new variants are introduced into development as 'terminal additions' (Gould's [1977] term) and then passed on. The adaptedness of form is, on Haeckel's

view, not of any great theoretical significance [Bowler, 1988]. But here again is a theory in which the development of traits and their inheritance are consequences of the same processes of ontogeny. These theories founder, if they do, on empirical grounds and not because they conflate the processes of inheritance, development and adaptation.

But why should the modern synthesis require the fractionation of evolution? The modern synthesis theory, as usually conceived, stands on three conceptual pillars: Darwinism, Mendelism and Weissmanism. This foundation, I believe, not only fragments the constituent processes of evolution into three discrete processes, it also abets the marginalization of development.

2.1 Darwinism

Darwin may not have been the first to realize that biology's two great explananda — the fit of organisms to the conditions of their existence and the diversity of organic form — are the joint consequences of evolution. His great achievement was the discovery of what sort of process evolution is: it is a *population-level* process. Darwin taught us to ask not how organisms change in ways that bring about their adaptedness, but rather to ask how *populations* come to comprise organisms so wonderfully adapted. Darwin's answer, of course, is natural selection. Natural selection occurs in a population that exhibits heritable variation in fitness. Not just any variation in fitness will do; the variants must be familial (i.e. passed on from parent to offspring). Offspring must resemble their parents with respect to the traits that contribute to differential survival and reproduction if these are to spread throughout the population in subsequent generations.

While it requires heritable variation in fitness, Darwin's process makes very few demands on how variants are generated, or the mechanisms that secure their reliable recurrence from one generation to the next. True, certain processes of inheritance better promote the capacity of selection to drive change in a population. Selection destroys the variation it requires and a process of inheritance that tends to preserve variants is more propitious for the continuation of selection. But any mechanism that underwrites the resemblance of offspring to parents will suffice to some degree. Nor does selection constrain the nature of development in any particularly strong way. Any process by which inherited traits develop reliably will promote selection.[1] So there is nothing about the conceptual structure of natural selection theory that demands the fractionation of evolution into three discrete processes of inheritance, development an adaptation. After all, as is well known, Darwin was happy to endorse the unity of inheritance and development found in Lamarck. Still, under Darwinism development and inheritance are mostly isolated from the process of *adaptation*, as they make no significant contribution to the *adaptiveness* of variants. It is selection that causes adaptation.

[1] Of course, some mechanisms of development will be more conducive to the capacity of selection to promote *adaptive* evolution. More on this below.

2.2 Mendelism

Mendel's theory holds that inheritance is particulate. Even if there should be some blending of inherited variants within a generation, the Mendelian mechanism of inheritance allows the variant factors to be recovered and transmitted intact from generation to generation without change. Mendel's process was made to measure for Darwin's theory.[2] When combined with the Mendelian theory of inheritance, the theory of selection tells us that adaptive evolution is the change in relative frequency of *unchanging* inheritable factors within a population, as a function of their systematic contribution to differential survival and reproduction. Because the factors remain unchanged through successive iterations of selection, variation — the grist for selection — is preserved in the population for longer than it would be under blending inheritance, and this enhances the efficacy of selection [Fisher 1930, Ch 1]. When Darwinism is combined with Mendelism, invariant factors become the units of inheritance, and evolution is a change in their relative frequencies.

Even so, selection has an overall tendency to eliminate variants over time. Darwin's process needs a source of new variants and Mendelism, in it purest form, is non-committal about that source. It is consistent with Mendelism that new variants may be introduced into a population through the process of development, and then inherited as invariant factors.[3] Nevertheless, the Mendelian process generally leaves development little or no role in determining the content of those traits transmitted from parent to offspring.

2.3 Weissmannism

August Weissman's experimental work demonstrated two important facts about development. The first is that the germ plasm of animals is sequestered early in development. It is only elements of the germ plasm that are transmitted from parent to offspring in reproduction. Developmental differentiation takes place (almost) exclusively in the somatoplasm. The germ plasm is quarantined from changes wrought on the somatoplasm during an organism's development. There is no interchange of materials between these two lineages of cells, and very little influence of one on the other. The second important fact is that at least *some* changes in organismal form introduced into the somatoplasm fail to be passed on to the following generation. Together these observations have been widely taken to entail that the only heritable traits are those whose development is 'encoded' in the germ plasm at an organism's inception. Given that, the new variants on which selection works must generated exclusively from *within* the heritable material. Development thus has no bearing on what is inherited.

The Weissman doctrine further peripheralizes the process of development from the explanation of fit and diversity. The fit and diversity of organisms comes about through the process of the generation of novel variants in the heritable material,

[2]The difficulties in getting the initial fit right, notwithstanding [Provine, 1971].

[3]This is a conceivable position in logical space, although I'm unaware of any early 20^{th} century thinker who occupies it.

their retention from one generation to the next and their subsequent selection. The iterated cycle of inheritance and selection requires that the processes of ontogeny transmit the heritable variants reliably. But, apart from that, ontogeny plays no significant role in the creating or fixing either the diversity of organisms or generating their adaptedness.

So we have three distinct processes, development, inheritance and adaptation, each with a significant degree of autonomy. Adaptive evolution requires all three. Without selection, there is no evolutionary change. But selection requires the reliable recurrence (inheritance) of traits, and a source of new variants. The process of Mendelian inheritance meets these requirements. But variants are selected through their contribution to the survival and reproduction of entire organisms. Ontogeny is the process that delivers Mendelian factors, embodied in organisms, to the arena of selection.

3 SUB-ORGANISMAL BIOLOGY

The fragmentation of evolution into three independent processes also calls for an account of what unites them into a single process of adaptive evolution. The concept of the replicator plays this unifying role in orthodox versions of current biology. Replicators are the only units of biological organization that take part in all the constituent processes of evolution, and whose causal capacities unite these processes into the process of evolution. (i) *Inheritance*: Replicators, we are told, are the units of inheritance. They are the only entities that are both copied within parents and transmitted from parent to offspring in reproduction. (ii) *Ontogeny*: Development, according to the sub-organismal view, is simply the implementation of a developmental program, the decoding of phenotypic information encoded, in the replicators (e.g. [Maynard Smith, 2000]). Organisms may be what are selected, but it is replicators that build the organisms that are selected. (iii) *Adaptation*: Adaptive evolution by natural selection occurs when populations change their structure as a consequence of the differential survival and reproduction of organisms. But not just any old change in population structure counts as evolutionary change — even a change that increases the average adaptedness of individuals in the population — adaptive evolution occurs, we are told, only when the relative frequency of replicators change. Indeed natural selection is usually *defined* as changes in replicator frequencies that result from the differential contribution of replicators to survival and reproduction [Bell, 2000]. Replicators are not only the causes of evolution, they are its currency.

The most thorough and provocative exponent of sub-organismal biology is Richard Dawkins:

> Evolution is the external and visible manifestation of the survival of alternative replicators... Genes are replicators; organisms... are best not regarded as replicators; they are vehicles in which replicators travel about. Replicator selection is the process by which some replicators

> survive at the expense of others. Vehicle selection is the process by which some vehicles are more successful than others in ensuring the survival of their replicators [1982, 82].

Dawkins serves up a particularly strongly distilled version of replicator biology, but it is the same in spirit as almost any orthodox account of evolutionary explanation. Perhaps replicator biology deserves its privileged status. It provides us with a powerful explanation of biology's dual problematic — fit and diversity. The marvellous fit of organisms to the conditions of their existence is achieved by the gradual accretion of selected (adapted) replicators. Fitness-promoting replicators are assembled and recombined by inheritance and selection into integrated suites of traits [Ayala, 1970].

The success of replicator biology has allowed evolutionists not merely to 'black box' development, but to see right through it; the organism becomes diaphanous. Dawkins drives home the point with great élan; watching through his window as a large willow tree sheds its seeds, he proclaims...

> It is raining DNA outside.... Up and down the canal, the water is white with cottony flecks.... The cotton wool is mostly made of cellulose, and it dwarfs the tiny capsule of that contains the DNA, the genetic information. The DNA content must be a small proportion of the total, so why did I say it was raining DNA rather than cellulose? The answer is that it is the DNA that matters.... The whole performance, cotton wool, catkins, tree and all, is in aid of one thing and one thing only, the spreading of DNA around the countryside. Not just any DNA, but DNA whose coded characters spell out specific instructions for building willow trees that will shed a new generation of downy seeds. Those fluffy specks are literally spreading instructions for making themselves. They are there because their ancestors succeeded in doing the same. It is raining instructions out there; it's raining tree-growing, fluff-spreading algorithms. That is not a metaphor, it is the plain truth. [Dawkins, 1987, 111]

Evolutionary biology is not about organisms, it would seem. Where organisms are of such minimal theoretical significance, their development is too. Development is merely an intra-organismal process. What really matters to evolution are the *inter*-organismal process of transmission of replicators and the *supra*-organismal process of selection. Small wonder, then, that as Hamburger notes, developmental biology was left out of the modern synthesis.

No one, not even the most ardent acolyte of the modern synthesis, denies that development has an explanatory role to play in evolutionary biology. Quite what that role might be, however, is a matter of some dispute. There are various attitudes we might take to the significance of development to the explanation of adaptive evolution. I attempt to outline a range of them in the next section. As I see it, it is not just the place of development in the sub-organismal reading of the

modern synthesis theory that is up for grabs here. If the most plausible account of the explanatory role of development is inconsistent with the fragmentational, sub-organismal interpretation of the modern synthesis, then so much the worse for that interpretation.

4 THREE GRADES OF ONTOGENETIC INVOLVEMENT

I distinguish three possible positions regarding the role of development in evolution. While these positions are probably neither exhaustive or mutually wholly exclusive, they do at least serve to demarcate increasing grades of commitment toward the significance of development to the study of evolution. As we move from one grade to the next, ontogeny takes on a more significant role in explaining the mechanisms of adaptive evolution. Only the first I believe is wholly consistent with replicator biology.

4.1 Grade I: Development as Constraint

Lewontin [1974] illustrates the relation between the transmission of replicators, ontogeny and the selection of organisms as the sequential interplay between two distinct property spaces: genotype space and phenotype space. The processes of transmission of inherited material from one generation to the next, the recombination of replicators, and the introduction of new mutations, induce changes in what Lewontin calls 'genotype space'. Their immediate effects are registered as changes in the relative frequencies of, or the relations between, genes. Selection, however, distinguishes between organisms on the basis of their phenotypes, consequently, its immediate effects are changes in 'phenotype space'. If evolution consists of changes in relative frequencies of genotypes as a function of selection, then it must involve an interplay between processes operating over genotype space and those operating over phenotype space. We need an account of the relation between these two spaces. It's easy to map changes in phenotype space onto changes in genotype space. When organisms differentially survive, reproduce or die, so do their genes. So changes in phenotype space are directly transposed into changes in genotype space. However, the mapping of changes in genotype space onto changes in phenotype space is less straightforward, because it is mediated by development. Development provides the so-called 'genotype/phenotype' map.

Organisms face the tribunal of natural selection as corporate entities, not as loose aggregates of traits. This fact imposes certain demands on development. At each stage development is conditioned by twin constraints: (i) it must maintain the organism's viability and (ii) it can only build phenotypes out of the materials and resources at its disposal. These demands introduce a bias into the range of available phenotypes that development can attain [Amundson, 1994]. Some potentially viable phenotypes may be closed off to development, because of the particular path it would have to take to attain them. Others may be particularly easy to attain, irrespective of the starting point in genotype space. Development biases the

mapping of genotype space into phenotype space. Some changes in genotype space may correspond to no changes in phenotype space. Small changes in genotype space may issue in large changes of phenotype space. This phenomenon goes by the name of 'developmental constraint'. Developmental constraint is defined as:

> ... a bias in the production of variant phenotypes or a limit on phenotypic variability caused by the structure, character, composition, or dynamics of the developmental system. [Maynard Smith *et al.*, 1985, 266]

C. H. Waddington [1942; 1957] was perhaps the first to make clear the significance of developmental constraint for evolution. Waddington drew attention to three salient features of development. The first is that an organism is made up of relatively few discrete, stable tissue types, all of which differentiate from a single, homogeneous precursor. The second is that the development of phenotype is heavily buffered. In a wild population, there is little phenotypic variation, despite an enormous amount of genetic variation and environmental perturbation.[4] The third is that given sufficient environmental stress, development can be induced to produce new stable phenotypes without a corresponding change in underlying genotype.

Quite what the implications of this might be for evolutionary biology has been an issue of some genuine concern and confusion ([Amundson, 1994]; [Maynard Smith *et al.*, 1985]; [Gould, 2002]). The most common approach is to suppose that while selection is the only adaptation-promoting force in evolution, it is to some degree at the mercy of ontogeny [Goodwin, 1982]. Sometimes development makes unavailable those phenotypes that, otherwise, selection would promote. Those who think of development playing primarily this role cast it as a constraint against the adaptation-promoting powers of selection.

> For instance developmental constraints frustrate selection by restricting the phenotypic variation selection has to act upon. Adaptations would be able to evolve only to optima within the constrained space of variability. [Wagner and Altenberg, 1996, 973]

> The nature of the existing developmental system somehow constrains or channels acceptable change, so that selection is limited in what it can achieve given some starting anatomy. [Raff, 1996, 294–295]

It doesn't follow from this that developmental constraint should play only a minor explanatory role. But it does follow that development and selection compete for explanatory relevance [Walsh, 2003]. Development and selection are both candidate explanantia of the distribution of biological form. Whereas selection distributes form in phenotype space by promoting *adaptedness*, development distributes form in a way that is *independent* of its adaptedness. Adaptedness is the

[4]The extent of the genetic variation was only realized with the advent of electrophoresis [Lewontin, 1974].

mark of selection. So development explains the distribution of form, only insofar as it is *non*-adaptive. Those who wish to accord a significant role to development in the explanation of biological form often find themselves in the uncomfortable position of having to deny what to many seems self-evident, that so much of biological form is adaptive.[5] The debate as it is played about between adaptationism and its adversaries often takes this form ([Gould and Lewontin, 1979]; [Sober, 1996]).

This conception of the role of development in evolution is what I call the 'First Grade of Ontogenetic Involvement'. I believe it is the orthodox position on the relevance of ontogeny for the process of evolution. First grade ontogenetic involvement is predicated upon the tripartite distinction among evolutionary processes handed down from the sub-organismal version of the modern synthesis. Development plays no role in the inheritance, recombination or initiation of variants in genotype space. Nor can it alter the adaptedness of populations, by driving a population up an adaptive peak in phenotype space; that is selection's job. Given the presumed fragmentation of the component processes of evolution, the only contribution developmental processes *could* make to evolution is that of a constraint against the adaptation-promoting power of selection.

There is ample reason to resist the fragmentation of evolution.[6] For one thing, there is no particularly compelling reason to suppose that replicators or genes are exclusively the units of inheritance. The principal motivation for this conviction resides in what Matteo Mameli [2005] calls the 'conception-donation' model of inheritance. On this view whatever is inherited is 'donated' by one parent or another at conception. Clearly, everything donated to an offspring at conception is a replicant derived from a parent.[7] But there is little reason to suppose that only these things can be inherited.

As we have seen, all selection requires of inheritance is the reliable recurrence of phenotypes within familial lineages. Replication may well guarantee this, but there is little reason to think that *only* replication does. There are myriad mechanisms by which such reliable recurrence may be secured. Variant phenotypes can be passed on by imprinting [Mameli, 2004], learning, cultural transmission ([Jablonka and Lamb, 2004]; [Boyd and Richerson, 2005]), the constancy of ecological conditions [Gilbert, 2001] and crucially any number of epigenetic and developmental mechanisms [Jablonka, 2001]. Replication may be inheritance, but inheritance is not replication.

This point is elegantly illustrated by a parable offered by Mameli [2004]. A population of butterflies lays its eggs on a certain species of host plant. The

[5]The adaptedness of form seems so self-evident (at least to some) that it inspired John Maynard Smith to declare "The main task of any evolutionary theory is to explain adaptive complexity, i.e. to explain the same set of facts which Paley used as evidence of a Creator [Maynard Smith, 1969].

[6]Notwithstanding the gifts of the three 'Weissmen' discussed above.

[7]I take it that anything that is copied and passed on is a replicant; this category includes more than just genes. Methylation patterns, maternal cytoplasm, cell membranes [Jablonka and Lamb, 2005] all qualify on this definition.

caterpillars imprint on their hosts; they lay their eggs on the type of plant on which they hatch. Suppose an egg on the host plant gets blown onto another type of plant, one that has until now never served as a host. This plant, as it turns out, is a beneficial host and the caterpillar that hatches on it has a considerable increase in fitness on account of this alone. There is now variation in fitness in this population, and it is heritable, as our new 'lucky' butterfly will lay her eggs on the newly encountered host. Because of its intergenerational stability, this variant phenotype is selectable. Adaptive evolution can occur in which this trait increases in prevalence because of its contribution to survival and reproduction. Yet *ex hypothesi* there is no change or variation in genetic material and no replication of the inherited trait. This form of inheritance is not inconsistent with the Mendelian mechanism — but it doesn't instantiate it — nor does it violate the Weissman doctrine of sequestration. Nevertheless, we have a change in a developmental process that is inheritable, selectable and adaptive. Mendelism and Weissmanism are sort of an irrelevance.

This suggests that the fragmentation of evolution is in part mistaken. There may well be discernible Mendelian, Darwinian and Weissmanian (developmental) processes, but it doesn't follow that inheritance and the generation of selectable variants is the domain of the Mendelian process alone. Development can generate and preserve selectable variants too. This insight motivates the second grade of ontogenetic involvement.

4.2 Grade II: Developmental Processes as Selecta

The lucky butterfly scenario brings into relief the importance of a biological truism: phenotypes are not the consequence of replicators alone. Genes may exert some control over phenotype, but they do not do so in glorious isolation; it takes genes and environments to make a phenotype. The environment constitutes a resource for the developing organism. This isn't news, of course, and biologists have long grown accustomed to thinking that despite the truism there is a still a privileged role for replicators to play in the production of phenotypes. Replicators are crucial difference makers. They are presumed to play a more significant, or perhaps more reliable, role than environmental factors do in determining that an organism will have one phenotype rather than some other. But replicators deserve no such deference. Advocates of ecological developmental biology stress that ecological conditions are frequently just as important in the determination of specific phenotypes [Gilbert, 2001]. Sex-determination in reptiles is often triggered environmentally. Polyphenisms are induced by specific environmental cues. Even heritable traits show this dependence on ecological factors. The pattern of deposition of bone in vertebrates is both highly heritable and highly sensitive to the gravitational attraction of the earth. Even heritable traits that differ systematically between lineages may be under the control of ecological conditions (e.g. the lucky butterfly). This is not to say that replicators aren't crucial difference makers for inheritable traits; it is simply to insist that in principle, any develop-

mental resource can be, whether it be a genetic, epigenetic or ecological factor. The contribution that replicators make to phenotype is no different in kind than that of any other developmental resource.

This is the central and important insight of Developmental Systems Theory (DST) ([Oyama, 1985]; [Oyama *et al.*, 2001]). DST encompasses a cluster of theses about the relation of inheritance and development. In contrast to replicator theory, it holds that the fundamental units of inheritance and phenotypic control are not genes or replicators, but developmental systems. A developmental system comprises the complete set of resources — genetic, epigenetic, ecological — required for the development of a trait (or organism). Control of the phenotype is distributed throughout the organism/environment system.The basic unit of inheritance, then, is that set of resources, or affordances, sufficient to secure the reliable intergenerational recurrence of a phenotype.[8]

The First Grade of ontogenetic involvement insists that inheritance and development are wholly distinct processes. There is, consequently, a corresponding distinction between traits that are inherited and those that are merely acquired during development. The developmental acquisition of traits does not impinge on the intrinsic structure of the replicators (or their 'informational' content).[9] DST denies the distinction between inherited (innate) and developmentally acquired traits, because it denies that developmental processes can be excluded from the mechanisms of inheritance.[10] There is no principled distinction between the processes that secure inheritance of traits and those that secure the development of individuals. The reliable recurrence of traits across generations enlists the full panoply of developmental resources. The schism drawn by sub-organismal, replicator biology between the processes of inheritance and development is illicit. Inheritance is simply the reliable transgenerational recurrence of developmental processes. Development is the key to understanding it. I call this conception of the role of development in evolution, 'Grade II Ontogenetic Involvement'.

Again, the influence of C. H. Waddington is apparent in the second grade of ontogenetic involvement. Waddington [1942; 1957] was perhaps the first to articulate the idea that we should see the variants among which selection selects as processes rather than static traits. For Waddington, a principal message of canalization, the epigenetic control of development, is that developmental processes — not just organisms — change during adaptive evolution. Canalization lowers the threshold for the development traits, or entrenches them (or both), rendering their development more easily attained, or more robust against genetic and environmental perturbations (or both). The consequences of adaptive evolution

[8]Griffiths and Gray stress that DST "... applies the concept of inheritance to any resource that is reliably present in successive generations, and is part of the explanation of why each generation resembles the last". [2001, 196]

[9]This is the legacy of the Weissman doctrine, nowadays usually expressed as the 'central dogma'.

[10]Some authors, notably Ariew [1996; 1998, this volume] insist that nevertheless a theoretically useful distinction (albeit one of degree) between innate and acquired traits can be salvaged. Innate traits are canalized.

are not simply measured in changes in genotypes, but changes in developmental processes. We should measure evolution, then, not as change of relative gene frequencies, but change in relative frequencies of inheritable developmental systems. Grade II ontogenetic involvement, then, stands in stark contrast to first Grade involvement.

Yet Grade II ontogenetic commitment does preserve at least some of the distinctions among evolutionary processes that mark out Grade I. It preserves the independence of the supra-organismal process of adaptation from the organism-level processes of inheritance and development. Adaptation, on this view, just as in Grade I, is the consequence of selection operating over heritable variants [Griffiths and Gray, 2001; 2004]. Indeed advocates of DST take great pains to emphasize that their approach to development and inheritance leaves the integrity of the Darwinian process untouched.[11]

> We hope it is now clear how DST can explain adaptation, in the modern sense of that term. Change over time in the developmental system of a lineage is driven by the differing capacity of variant developmental systems to reconstruct themselves, or, in a word, differential fitness. What is fitness? In contemporary evolutionary theory fitness is a measure of the capacity of a developmental system to reproduce itself.... [2001, 209]

In their conclusion to the same article Griffiths and Gray offer some definitions of standard evolutionary terms from the perspective of DST:

> Natural Selection — the differential reproduction of heritable variants of developmental systems due to relative improvements in their functioning.
>
> Adaptation — the product of natural selection. [Griffiths and Gray, 2001, 214]

Grade II commitment takes a conservative approach to the adaptiveness of adaptive evolution. Explaining the adaptedness of organisms to the conditions of their existence is the exclusive domain of the population-level process of selection. The individual-level processes of inheritance and development make no great contribution to adaptedness, except insofar as they serve up the variants among which selection winnows. Grade II involvement substitutes developmental systems for genes as units of selection [Griffiths and Gray, 2004], but its conception of the process of adaptation is still resolutely sub-organismal.

The charge that DST is 'sub-organismal' is bound to rankle its adherents. After all, it as a DST shibboleth that the organism/environment boundary is arbitrary or irrelevant ([Oyama, 1985]; [Oyama *et al.*, 2001]). DST is to be credited with the recognition that developmental systems extend beyond the skin, but that alone

[11] Depew and Weber [2001] have their doubts.

doesn't make DST's approach to adaptation non-sub-organismal. This is a phenomenon that replicator biology is fully able to accommodate. The concept of the extended phenotype was introduced in acknowledgement of the fact that an organism's traits extend beyond the skin. The concept of the extended phenotype contends that replicators contain information for building features of an organism's environment (e.g. birds' nests, beehives and beaver dams) [Sterelny, 2001]. The recognition that inheritable phenotypes can exist beyond thee skin is not sufficient to make for a non-sub-organismal biology.

DST is sub-organismal with respect to its approach to adaptive evolution. It falls within the ambit of the DST approach to explain the adaptedness of organisms. And just as in Grade I involvement, the process by which organisms become adapted is one of selection, recombination and reassortment among a population of units of inheritance. The adaptedness of organisms is the consequence of the generation, recombination, reassortment, and selection, of these entities. These selecta are not organisms, but developmental systems, aggregates of which together constitute an organism.[12] As in Grade I, Grade II adaptation occurs through the accretion within organisms of adaptive (selected) entities. Selection alone is what makes adaptive evolution adaptive. So adaptedness of organisms is explained by the aggregation of sub-organismal selected entities.

The adequacy of selection alone as a cause of the adaptedness of organisms has recently been called into question. Not just any lineage of reproducing entities can increase in adaptiveness under a process of mutation and selection. This has been one of the central findings of research into genetic algorithms. A genetic algorithm must have certain intrinsic, structural properties if it is to undergo adaptive evolution [Wagner and Altenberg, 1996]. One of the projects of genetic algorithm research is to address the question, what sorts of properties must programmes have in order that they may undergo *adaptive* evolution. Biologists have recently adapted that question to the context of evolutionary biology under the rubric of 'evolvability'. Evolvability is defined as "... an organism's capacity to generate heritable phenotypic variation" [Kirschner and Gerhart, 1998]. The question is 'What properties must organisms have such that lineages of them may undergo adaptive evolution?'[13] Schwenk and Wagner outline the basic requirements.

> On the one hand, phenotypes must be mutable and therefore responsive to the constantly changing demands of the environment.... On the other hand, phenotypes must be stable so that the complex dynamics of their developmental and functional systems are not disrupted.... It is this fundamental tension — between mutability and stability — that current evolutionary biology seeks to explain. [2003, 390]

But how are stability and mutability achieved? Recent developmental biology suggests that stability and mutability are both consequences of a fundamental,

[12] At least aggregates of their *phenotypic effects* constitute an organism.

[13] Kirschner and Gerhart [1998], Von Dassow and Munro [1999], Wagner and Altenberg [1996].indexWagner, G.

defining property of whole organisms, a property that is most clearly manifested in development. This brings us to our third grade of ontogenetic involvement.

4.3 Grade III: Development as Adaptive

The distinguishing feature of Third Grade ontogenetic involvement is its rejection of sub-organismal biology. In Grades I and II the distinctive capacities of organisms play no significant role in explaining inheritance, development and adaptation. In Grade III the capacities of organisms explain and unify all three processes. Evelyn Fox Keller characterizes an organism in the following way:

> It is a bounded physico-chemical body capable not only of self-regulation — self-steering — but also, and perhaps most important, of self-forming. An organism is a material entity that is transformed into a self-generating "self" by virtue of its peculiar and particular organization. [Fox Keller, 2000, 108]

Organisms are adaptive entities, capable of attaining their developmental end states, and maintaining their viability in the face of all manner of vagaries of their environment, genome and developmental systems [Gibson, 2002]. They are the very paradigm of a robust self-organizing system ([Kitano, 2004]; [Goodwin *et al.*, 1993]).

The definitive feature of organisms is phenotypic plasticity, 'a universal property of living things' [West-Eberhard, 2003]. Plasticity is '...the ability of an organism to react to an internal or external environmental input with a change in form, state, movement, or rate of activity' [West-Eberhard, 2003, 33]. Recent work in development suggests that phenotypic plasticity confers on organisms both stability and mutability. The stability of organisms is achieved through their capacity to make compensatory changes in their physiology and development to perturbations. Plasticity also endows organisms with the capacity to respond to unpredictable circumstances through the initiation of entirely novel, yet stable phenotypes. Plasticity, then, fulfils *both* conditions for evolvability.

Mary Jane West-Eberhard has proposed that adaptive evolution, and diversification, are driven primarily by the control that the plasticity of the phenotype exerts over the constituent processes of an organism's development. Adaptive evolution proceeds in three stages stages (i) phenotypic accommodation, (ii) developmental recombination and (iii) genetic accommodation, all of which depend upon the plasticity of organisms.[14]

> *Phenotypic accommodation*: Plasticity allows organisms to accommodate to the features of their environment (and their genomes) in ways that preserve their well functioning, by making subtle compensatory adjustments to their phenotypes. "Phenotypic accommodation due to

[14] In her 2003 West-Eberhard explicitly sites two stages; in her 2005b paper there appear to be three.

> phenotypic plasticity is the immediate adaptive adjustment of phenotype to the production of a novel trait or trait combination . . . Phenotypic accommodation reduces the amount of functional disruption occasioned by developmental novelty." [West-Eberhard, 2003, 147]

> "Adaptive phenotypic adjustment to potentially disruptive effects of the novel input exaggerate and accommodate the phenotypic change *without genetic change*." [West-Eberhard, 2005a, 613].

The accommodation of one part of an organism's phenotype often requires, and elicits, accommodation by others. In the process it occasionally produces novel, stable, adaptively advantageous phenotypes. The generation, by development, of novel adaptive phenotypes has further consequences.

> *Developmental recombination*: "Developmental recombination occurs in a population of individuals because of a new, or recurrent input. A new input... causes a reorganization of the phenotype, or "developmental recombination". Given the variable developmental plasticity of different individuals, this process produces a population of novel variable phenotypes, providing material for selection". [West-Eberhard, 2005b, 6544].

For any novel phenotype, there may be within the population many alternative developmental systems capable of producing it; some combinations of them will be better than others at reliably producing the new phenotype. Repeated developmental recombination will have the effect of lowering the threshold for the development of an adaptively advantageous novelty (*Cf* [Waddington, 1957]).

> *Genetic accommodation*: "If the phenotypic variation is associated with variation in reproductive success, natural selection results; and to the degree that the variants acted upon by selection are genetically variable, selection will produce genetic accommodation..." [West-Eberhard, 2005a, 612]. Genetic accommodation is "...adaptive evolution that involves gene-frequency change". [West-Eberhard, 2005b, 6544].

On this model, adaptive plasticity, a feature of development, is both the cause of the adaptedness of organisms and the diversity of organic form.

This is a radical proposal. It reverses the causal priority of genotype over phenotype in evolution that is the cornerstone of sub-organismal, replicator interpretation of the modern synthesis. Phenotypic novelties are initiated in development and not by mutation. Palmer [2004] calls this a "phenotype before genotype" approach to novelty. As for novelties, so for changes; evolutionary changes in phenotype space drives changes in genotype space: "genes are followers in evolution, not leaders" [West-Eberhard, 2003].

The phenotype first model needs an account how the mechanisms of development generate adaptive phenotypic novelties. Recent research into developmental biology is beginning to reveal these very mechanisms. The plasticity of organisms is a consequence of the modular architecture of development. Developmental modules are internally integrated, mutually dissociated, units of developmental control.

> In principle, a modular system allows the generation of diverse phenotypes by the rearrangement of its internal module connections and relatively independent evolution of each module... [Kitano, 2004, 830]

The internal structure of a module typically consists of a series of integrated regulatory feedback relations. This confers on a module both robustness — the capacity to function reliably across a range of perturbations ([von Dassow and Munro, 1999]; [Kitano, 2004]) — and flexibility, the capacity to mount novel adaptive changes to perturbations ([Greenspan, 2002]; [Bergman and Siegal, 2002]). The dissociation of modules — the fact that each influences only a few others — has two distinct kinds of effects. The first is that a module is isolated from potentially deleterious changes in other parts of the developing organism. The second is that the function of a module can be stably varied, according to its context. The repertoire of a developmental module, the range of its capacities taken in isolation, is considerably greater than the range of its realized effects [Greenspan, 2002]. Typically each module is capable of producing any number of a large array of stable outputs [Von Dassow *et al.*, 2000]. Which of its capacities is manifested on a particular occasion is determined by the context in which the module finds itself. Taken together, these properties of modular architecture confer on an organism the capacity robustly and reliably to produce a viable organism typical of its kind, *and* to generate phenotypic novelties as an adaptive response to genetic, epigenetic or environmental perturbations. These twin capacities are crucial to adaptive evolution.

Kirschner and Gerhard [2005] propose that the process of adaptive evolution proceeds through the 'facilitated variation'. An organism's development is built on the foundation of a suite of highly robust 'conserved core processes'. These are fixed in their function and are significantly immune to perturbation. These core conserved processes 'deconstrain' other component processes of development. That is to say, they underwrite the capacity of other component processes of development to produce adaptive changes. *Hox* gene clusters, for example, are highly constrained core processes. They orchestrate the development of an enormously wide range of morphological structures. The same *Hox* clusters regulate the development of vastly different structures in different organisms. And they initiate the development of widely varying structures within the same organism. The duplication of *Hox* clusters, their change of timing, or context can bring about remarkable phenotypic changes. Carroll *et al.* [2002] conjecture that most of the large scale changes in metazoan morphology have been caused by changes in the regulatory roles of core conserved elements like *Hox* clusters.

Kirschner and Gerhart concur:

> Most evolutionary change in the metazoa since the Cambrian has come not from changes of the core processes themselves, but from regulatory changes affecting the deployment of the core processes. ... Because of these regulatory changes, the core processes are used in new combinations and amounts at new times and places. [Kirschner and Gerhart, 2005, 221–222]

West-Eberhard summarizes the view succinctly. "A very large body of evidence ... shows that phenotypic novelty is largely reorganizational rather than a product of innovative genes" [2005b, 6547]. Because phenotypic novelties are the result of an organism's capacity to make adaptive responses to change, the novelties are not random. They tend to be adaptive: "Variation ... is both less lethal and more appropriate to selective conditions than would be variation from random change. Evolutionary change is thereby facilitated" [Kirschner and Gerhard, 2005, 221].

There are two important points to be made about the contribution of phenotypic plasticity to adaptive evolution: (i) phenotypic plasticity generates phenotypic novelty, and (ii) the phenotypic novelty it produces is not adaptively neutral. We must recognize, then, that development has a central place amongst the causes of adaptive evolution. It initiates adaptive novelties; it secures the conditions for adaptive change in populations; it supplies adaptive novelties within a population; and it provides a means for making the development of adaptive novelties robust and routine. On this view, it is the plasticity of development that makes adaptive evolution adaptive. "This analysis brings development, largely omitted from evolutionary biology during the synthesis era ..., to the forefront of evolutionary biology as the source of the variation that fuels natural selection and adaptive evolution." [West Eberhard, 2005b, 6544]

This is what I call the 'Third Grade of Ontogenetic Involvement'. This model of adaptive evolution confers a role on development that neither Grade I nor Grade II involvement does. In grades I and II, development plays no role in making evolution adaptive. The sole adaptation promoting mechanism is natural selection. Grade III makes no distinction between the organism-level processes of inheritance and development and the population-level causes of adaptation. The Grade III maxim: *evolution is adaptive because ontogeny is adaptive.*

In fact, Grade III involvement, emphasizing as it does the importance of processes internal to organisms, suggests a particular 'ontogeny first' metaphysics of evolution. Organisms develop and reproduce. The plasticity of development secures the resemblance of offspring to parents — inheritance — despite the (internal and external) vagaries of the world. So inheritance is the simple consequence of the plasticity of development plus reproduction. The plasticity of development also generates novel adaptive phenotypes. Through reproduction, developmental processes that produce these novel phenotypes, are assorted and recombined. This lowers the threshold for the occurrence of these adaptive novelties within individuals and causes their spread throughout the population. So ontogeny causes

adaptive evolution too.

The ontogeny-first conception of evolution does not deny the existence of Darwin's process. Selection, as it were, occurs spontaneously. One doesn't need to add a distinct population-level process to the mix to get adaptive evolution to occur. We have all the ingredients we need in the development and reproduction of individual organisms. Nor does one need to posit an independent, supra-organismal causes of the adaptiveness of evolutionary change. Selection is a genuine process, of course, perhaps even a causal process, but it doesn't contribute any causes to adaptive evolution that aren't already fully accounted for by the development and reproduction of individual organisms. The population-level Darwinian process — even the increased in average adaptedness of organisms in a population — is just the resultant of the suite of individual-level processes of development and reproduction. The sum of these processes is *adaptive* evolution, because of the adaptive capacities of organismal development.

On Grade III ontological commitment, the plasticity of development (plus some reproduction) causes: (i) the reliable recurrence of traits from generation to generation (inheritance) (ii) the origin of adaptive novelties and (iii) the increase in the adaptedness of organisms in a population to their conditions of existence (adaptation). Ontogeny is the central unifying process in all of evolution.

5 CONCLUSION: GOING THROUGH THE GRADES

I have attempted to outline a space of alternative views on the significance of organismal development for the explanation of adaptive evolution. For the most part biologists, philosophers of biology, and most certainly non-specialists in biology labour under a conception evolution forged in the early days of the modern synthesis, and then tempered by generations of selectionism. On this account, development is of minor, sometimes negligible significance. There may be good sociological, historical, even technological explanations of the marginalization of development; perhaps it is harder to study development in detail than much of paleontology, or ethology, or ecology, or systematics, or population genetics, or any of the disciplines that coalesced into the synthesis. Certainly the current renaissance of developmental biology had to await a raft of technological advances in molecular biology. Nevertheless, it is important to acknowledge that there are conceptual reasons too. The fact that ontogeny plays only a peripheral role in evolutionary biology is due to a commitment embodied in the predominant version of modern synthesis biology to the fragmentation of evolution into three discrete, largely autonomous processes: development, inheritance, and adaptation. This fragmentation has consequences for the place of development in evolutionary biology; it leads to the marginalization of development. But the fragmentation isn't obligatory.

I have outlined three grades of commitment to the significance of ontogeny to the explanation of adaptive evolution. As we progress through the grades (from Grade I to Grade III) we see two related trends (i) an increase in the significance

of development for the process of adaptive evolution and (ii) the reunification of evolution's constituent processes. In Grade I involvement development plays no role in either inheritance or in the production of the fit and diversity of organisms. Adaptive evolution and the divergence of lineages are the results of random mutation and transmission of inherited variants and their selection. This requires the recognition of three discrete processes; inheritance of replicators, development of organisms and adaptive change of populations, each with its proprietary causal process. Grade II involvement draws development more centrally into the process of evolution; the functioning of developmental systems — and not the replication of genes — explains inheritance; inheritance is simply the reliable transgenerational re-occurrence of developmental processes (systems). There is, then, no distinction between the processes that underwrite inheritance and those that underwrite development; they're both ontogeny. In Grade II, development makes no positive contribution to the adaptiveness and diversification of lineages. This is exclusively the domain of natural selection. Grade III involvement preserves the unity of the processes of inheritance and development from Grade II, but it casts ontogeny in an even more central role. According to Grade III, the plasticity of organismal development is the principal cause of adaptive evolution. Grade III involvement is consistent with natural selection, of course. But selection — adaptive evolution — is just a consequence of the development of individual organisms. The processes of development, inheritance and adaptation are all fundamentally just ontogeny.

As far as I can tell, these are all coherent positions. They are seldom, if ever, articulated in quite this way, but I think it's fair to say that, implicitly or explicitly, each has its adherents. Nor do these positions seem to be empirically equivalent; each has observational consequence that the others don't. Which of these positions is the most plausible, then, looks like an empirical issue. It is one that I believe developmental biology only now has the resources to address.

BIBLIOGRAPHY

[Amundson, 1994] R. Amundson. Two concepts of constraint: Adaptationism and the challenge from developmental biology. *Philosophy of Science*, 61: 556–578, 1994.

[Ariew, 1996] A. Ariew. Innateness and canalization. *Philosophy of Science*, 63 (Proceedings):S19–S27, 1996.

[Ariew, 1998] A. Ariew. Innateness is canalization: A defense of a developmental account of innateness. In V. Hardcastle (ed.), *Biolology Meets Psychology: Conjectures, Connections, Constraints*. Cambridge, Ma.: MIT Press, 1998.

[Ariew, forthcoming] A. Ariew. Triggers (this volume), forthcoming.

[Ayala, 1970] F. Ayala. Teleological explanations in biology. *Philosophy of Science*, 37: 1–15, 1970.

[Bell, 2000] G. Bell. *Selection: The mechanism of evolution*, New York: Chapman Hall, 2000.

[Bergman and Siegal, 2002] A. Bergman and M. Siegal. Evolutionary capacitance as a general feature of complex gene networks. *Nature*, 424: 549–552, 2002.

[Bowler, 1988] P. Bowler. *The Non-Darwinian Revolution*. Baltimore: Johns Hopkins University Press, 1988.

[Boyd and Richardson, 2004] R. Boyd and P. Richardson. *Not by Genes Alone*. Chicago; University of Chicago Press, 2004.

[Dawkins, 1982] R. Dawkins. *The Selfish Gene*. Oxford: Oxford University Press, 1982.

[Dawkins, 1986] R. Dawkins. *The Extended Phenotype*. Oxford: W.H. Freeman, 1986.
[Dawkins, 1987] R. Dawkins. *The Blind Watchmaker*. New York: W.W. Norton and Co., 1987.
[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. New York: Dover, 1930.
[Fox Keller, 2000] E. Fox Keller. *The Century of the Gene*. Cambridge, Ma.: Harvard University Press, 2000.
[Gibson, 2002] G. Gibson. Getting robust about robustness. *Current Biology*, 12: R347–9, 2002.
[Gilbert, 2001] S. Gilbert. Ecological developmental biology: Developmental biology meets the real world. *Developmental Biology*, 233: 1–12, 2001.
[Goodwin, 1982] B. Goodwin. Development and evolution. *Journal of Theoretical Biology*, 97: 43–55, 1982.
[Goodwin *et al.*, 1993] B. G. Goodwin, S. Kauffman, and J. D. Murray. Is morphogenesis an intrinsically robust process? *Journal of Theoretical Biology*, 163: 135–144, 1993.
[Gould, 1977] S. J. Gould. *Ontogeny and Phylogeny*. Cambridge, Ma.: Belknap Press, 1977.
[Gould, 1983] S. J. Gould. The hardening of the modern synthesis in Grene, M. *Dimensions of Darwinism*, Cambridge: Cambridge University Press, 1983.
[Gould, 2002] S. J. Gould. *The Structure of Evolutionary Theory*. Cambridge, Ma.: Harvard University Press, 2002.
[Gould and Lewontin, 1969] S. J. Gould and R. C. Lewontin. The Spandrels of San Marco: A critique of the Panglossian paradigm. *Proceedings of the Royal Society, London: B*, 1969.
[Griffiths and Gray, 2001] P. E. Griffiths and R. D. Gray. Darwinism and developmental systems. In S. Oyama, P. E. Griffiths, and R. Gray (eds.), *Cycles of Contingency*. Cambridge, Ma.: MIT Press, pages 195–218, 2001.
[Griffiths and Gray, 2004] P. E. Griffiths and R. D. Gray. The developmental systems perspective: Organism-environment systems as units of evolution. In K. Preston and M. Pigliucci (eds.), *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*. Oxford: Oxford University Press, pages 409–431, 2004.
[Hamburger, 1980] V. Hamburger. The evolutionary synthesis. In E. Mayr and W. Provine (eds.), *The Evolutionary Synthesis*. Cambridge Ma.: Cambridge University Press, pages 96–112, 1980.
[Jablonka, 2001] E. Jablonka. The systems of inheritance. In S. Oyama, P. E. Griffiths, and R. Gray (eds.), *Cycles of Contingency*. Cambridge, Ma.: MIT Press, pages 99–115, 2001.
[Jablonka and Lamb, 2005] E. Jablonka and M. Lamb. *Evolution in Four Dimensions*. Cambrdige, Ma.: MIT Press, 2005.
[Kirschner and Gerhart, 1998] M. Kirschner and J. C. Gerhart. Evolvability. *Proceedings of the National Academy of Sciences USA*, 95: 8420–8427, 1998.
[Kitano, 2004] H. Kitano. Biological robustness. *Nature Reviews Genetics*, 5: 826–837, 2004.
[Lewontin, 1974] R. C. Lewontin. *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press, 1974.
[Mameli, 2004] M. Mameli. Non-genetic selection and non-genetic inheritance. *British Journal for the Philosophy of Science*, 55: 35–71, 2004.
[Mameli, 2005] M. Mameli. The inheritance of features. *Biology and Philosophy*, 20: 365–399, 2005.
[Maynard Smith, 1969] J. Maynard Smith. The status of neo-Darwinism. In C. H. Waddington (ed.), *Toward a Theoretical Biology*. Edinburgh: Edinburgh University Press, 1969.
[Maynard Smith, 2000] J. Maynard Smith. The concept of information in biology, *Philosophy of Science*, 67: 177–194, 2000.
[Maynard Smith *et al.*, 1985] J. Maynard Smith, R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert. Developmental constraints and evolution. *Quarterly Review of Biology*, 60: 265–287, 1985.
[Newman and Muller, 2001] S. Newman and G. Muller. Epigenetic mechanisms of character origination. In G. Wagner (ed.), *The Character Concept in Evolutionary Biology*, New York: Academic Press, pages 559–579, 2001.
[Oyama, 1985] S. Oyama. *The Ontogeny of Information*. Duke University Press, 1985.
[Oyama *et al.*, 2001] S. Oyama, P. E. Griffiths, and R. Gray. Introduction: What is Developmental Systems Theory? In S. Oyama, P. E. Griffiths, and R. Gray (eds.), *Cycles of Contingency*. Cambridge, Ma.: MIT Press, pages 1–12, 2001.
[Palemer, 20004] A. R. Palmer. Symmetry breaking and the evolution of development. *Science*, 306: 828–833, 2004.

[Provine, 1971] W. Provine. *The Origin of Theoretical populatrion Genetics*. Chicago: University of Chicago Press, 1971.

[Raff, 1996] R. Raff. *The Shape of Life: Genes, Development and the Evolution of Animal Form*. London: University of Chicago Press, 1996.

[Schwenk and Wagner, 2003] K. Schwenk and G. A. Wagner. The relativism of constraints on phenotypic evolution. In M. Pigliucci and K. Preston (eds.), *The Evolution of Complex Phenotypes*. Oxford University Press, pages 390–408, 2003.

[Sober, 1998] E. Sober. Six sayings about adaptationism. Reprinted in D. Hull, and M. Ruse (eds.), *The Philosophy of Biology*. Oxford: Oxford University Press, 1998.

[Von Dassow and Munro, 1999] G. Von Dassow and E. Munro. Modularity in animal development and evolution: Elements of a conceptual framework for EvoDevo. *Journal of Experimental Zoology (Mole Dev Evol)*, 285: 307–25, 1999.

[Von Dassow *et al.*, 2000] G. Von Dassow, E. Meir, E. M. Munro, and G. M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406: 188–92, 2000.

[Waddington, 1942] C. H. Waddington. Canalization of development and the inheritance of acquired characters. *Nature*, 150: 563–565, 1942.

[Waddington, 1957] C. H. Waddington. *The Strategy of the Genes*. London: Allen and Unwin, 1957.

[Waddington, 1960] C. H. Waddington. Evolutionary adaptation. In S. Tax (ed.), *Evolution after Darwin*. Chicago: University of Chicago Press, 381–402, 1960.

[Wagner, 2005] A. Wagner. *Robustness and Evolvability in Living Systems*. Princeton, N.J.: Princeton University Press, 2005.

[Wagner and Altenberg, 1996] G. Wagner and L. Altenberg. Complex adaptations and the evolution of evolvability. *Evolution*, 50: 967–976, 1996.

[Walsh, 2003] D. M. Walsh. Fit and diversity: Explaining adaptive evolution. *Philosophy of Science*, 70: 280–301, 2003.

[West-Eberhard, 2003] M. J. West-Eberhard. *Developmental Plasticity and Evolution*. Oxford: Oxford University Press, 2003.

[West-Eberhard, 2005a] M. J. West-Eberhard. Phenotypic accommodation: Adaptive innovation due to developmental plasticity. *Journal of Experimental Zoology (Mole Dev Evo)*, 304B:610–618, 2005a.

[West-Eberhard, 2005b] M. J. West-Eberhard. Developmental plasticity and the origin of species differences. *PNAS Sup.*, 102: 6543–6549, 2005b.

This page intentionally left blank

EVOLUTION AND NORMATIVITY

Michael Bradie

1 GENERAL INTRODUCTION

Human beings are complex organisms with highly developed physical attributes (e.g., upright walking and opposable thumbs) as well as highly developed emotional and intellectual capabilities. Evolutionary accounts of the origin and development of human beings offer a descriptive window into how we humans have come to be, on average, the creatures that we are. Among the intellectual qualities we possess are the ability to describe the world around us and the ability to communicate these descriptions to others. In addition to having the ability to say how things are, we have the ability to say how things *ought to be*. We are *pre*scribers as well as *de*scribers. The act of prescribing involves the articulation, endorsement and application of *norms*. Human beings are *normative beings*.

Norms are guides to how we ought to behave and what we ought to believe. Norms of behavior can roughly be classed as "ethical norms" although there is a range of cases here. At one end of the normative spectrum are items such as The Golden Rule and the Ten Commandments. At the other end are pragmatic rules of utility such as "Drive on the right (or left)." In between are a range of cases such as "Avoid incest," "Infanticide is wrong," "Abortion is wrong," and the like. Norms of belief can roughly be classified as "epistemic norms" and there is a range of cases here as well. Epistemic norms include rules of justification, evidential principles, principles of scientific methodology, and what might be called "engineering principles" of the form "If you want to build a better mousetrap, then . . .".

Providing examples of norms is fairly easy. All human societies and all human practices employ norms. The tricky questions come when one inquires about the *source* and *force* of norms. How do norms arise? Where do they come from? In certain cases the origin is clearly artificial or conventional. The rules of chess or of baseball, for example, are (low level) behavioral norms that we adopt and adhere to by convention. It seems less plausible to suggest that "serious" norms such as the Golden Rule are matters of artifice or convention. We are more inclined to believe that such rules are somehow rooted in our "natures." This opens up the possibility that, given that our "natures" are the product of evolutionary processes, we may somehow find some evolutionary grounding for these principles.

The second tricky question concerns the "force" of norms. Where does the force of norms come from and why should we obey them? Attempts to answer these

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

questions ultimately lead to the question of the justification of norms. The norms themselves justify our behavior or beliefs but what justifies the norms themselves? One traditional answer — we do it this way because we have always done it this way — falls afoul of the naturalistic fallacy. I do not propose to enter the tangled web of literature on the naturalistic fallacy and whether it is or is not a fallacy. Suffice it for our purposes that it strangely non-compelling to say that just because things *are* such and such a way that therefore they *ought* to be such and such a way. In any case, the naturalistic fallacy or rather the fear of committing it stands as a road block to the introduction of evolutionary facts and processes as considerations relevant to the justificational grounding of norms be they epistemic or ethical.

Another traditional answer is that moral norms are divine commands. This does not fare much better that the 'appeal to tradition' answer. James Rachels, for one, argued that Darwinism undermines the fundamental rationale for believing in ethics as grounded in divine commands [Rachels, 1990]. We return to this below. In any case, divine command theories run afoul of the Socratic dilemma as posed in the *Euthyphro*: Is something good or right because God commands it or does God command something because it is right or good? The former alternative can be seen as committing a version of the naturalistic fallacy. The latter alternative leaves the question of what makes something good or right open.

The Big Question then becomes: "Can our evolutionary history inform us in any meaningful way either about the source or the force of epistemic or ethical norms?" Philosophical opinion is here divided. On the one hand, there is the verdict of John Dewey writing fifty years after the appearance of Darwin's *Origin of Species*:

> Old ideas give way slowly; for they are more than abstract logical forms and categories. They are habits, predispositions, deeply ingrained attitudes of aversion and preference. Moreover, the conviction persists - though history shows it to be a hallucination - that all the questions that the human mind has asked are questions that can be answered in terms of the alternatives that the questions themselves present. But in fact intellectual progress usually occurs through sheer abandonment of questions together with both of the alternatives they assume - an abandonment that results from their decreasing vitality and a change of urgent interest. We do not solve them: we get over them. Old questions are solved by disappearing, evaporating, while new questions corresponding to the changed attitude of endeavor and preference take their place. Doubtless the greatest dissolvent in contemporary thought of old questions, the greatest precipitant of new methods, new intentions, new problems, is the one effected by the scientific revolution that found its climax in the "Origin of Species". [Dewey, 1910]

On the other hand, there is the quip by Ludwig Wittgenstein to the effect that "Darwin's theory has no more to do with philosophy than any other hypothesis in natural science" (*Tractatus*, 4.1122). Philosophical purists are liable to side with

Wittgenstein whereas pragmatists and naturalists are liable to side with Dewey. For my part, I am with Dewey in thinking that philosophers in general have not taken Darwin seriously enough and they ought to do so. Evolutionary theory and the facts of the evolutionary progress of human beings may not solve the big and deep philosophical questions but it is sheer parochialism to think that evolutionary considerations are entirely irrelevant to them.

In this paper I propose first to give an all too brief sketch of the major developments in evolutionary ethics and evolutionary epistemology. For a fuller treatment see [Bradie, 1986; 1994a; 1994b]. The central questions in evolutionary ethics and evolutionary epistemology have been addressed from a variety of perspectives. A failure to distinguish between them only serves to muddy further already unclear waters. In section 3, I shall propose and defend three distinctions that help to sort out the key issues. Finally, in section 4, I address the crucial question of what *substantive* contributions, if any, can evolutionary considerations bring to bear on the nature, the source and the force of epistemic and ethical norms.

2 THE SHADOW OF DARWIN

What can we learn about ethics or epistemology from the study of the evolution of human beings? Charles Darwin, who started it all, argued that an evolutionary account of what he called the 'moral sense' could explain the general shape of our capacity to make moral judgments [Darwin, 1981]. Following some leads from a relative by marriage, the philosopher James Mackintosh, Darwin sketched a plausible scenario to account for the development of the two poles of moral sentiment that serve to shape our moral judgments [Manier, 1978; Mackintosh, 1834]. The development of the proclivity for self-regarding or self interested motivation is driven by the needs of the organism to preserve itself in order to reproduce. The development of the proclivity for other-regarding motives or what we have come to label "altruism," is driven by evolutionary pressures generated by the social nature of human beings [Bradie, 1994b]. Darwin was less sanguine about the derivability of specific moral norms or values from the evolutionary history of human development.

Thomas Huxley, Darwin's 'bulldog,' famously argued that biological evolution and the development of ethics were antithetical. Biological evolution was driven, in his view, by competitive forces that were inimical to the development of values (Huxley 1989). Ethics, on the other hand, was based on co-operation and mutual aid. These qualities represented a triumph of human nature over the natural order of things. Nature, in his view, was morally neutral. Huxley's pessimistic vision was challenged by the Russian anarchist, Prince Peter Kropotkin, who argued that a proper reading of Darwinian theory represented natural selection as promoting both co-operative and competitive impulses. The net result was that Kropotkin saw the evolution of the moral sentiment of mutual aid as emerging from the evolution of natural structures (Kropotkin, 1919, 1924).

The latter half of the 19^{th} century saw a number of attempts to extend evolutionary insights to a wide range of human activities including politics, commerce, and social organization as well as morality (see e.g., [Spencer, 1966; Bagehot, 1956]). This attempt to read off social and moral norms from the 'face' of nature was met with skepticism and stiff resistance from philosophical quarters. Henry Sidgwick's *The Method of Ethics*, which first appeared in 1874, raised serious doubts about the extent to which natural circumstances could be said to give rise to moral norms [Sidgwick, 1964]. The philosophical *coup de gras* to deriving moral norms from the products of natural selection or biological evolution was delivered by G. E. Moore in 1905 who argued that systems such as Spencer's that drew moral consequences from natural facts suffered from a fatal flaw — commission of the 'naturalistic fallacy.' This was a 20^{th} century version of a warning from David Hume in the 18^{th} century about the legitimacy of deriving an "ought" from an "is" that has come to serve as the cornerstone of what is called the "fact/value distinction." Moore's version deplored the attempt to define "good," which Moore took to be unanalyzable and indefinable, in terms of 'natural' properties such as 'has evolved' or 'is an adaptation' [Moore, 1964]. This riposte, coupled with the rise of totalitarianism in the 1920's and 1930's and its claims to legitimacy grounded in appeals to Darwinian natural selection effectively muted attempts to ground ethics in evolutionary theory until the publication of E. O. Wilson's *Sociobiology* in 1975. There were, of course, some notable exceptions. Thus, Julian Huxley, Thomas Huxley's grandson, argued that his grandfather had been wrong –natural selection and the emergence of ethics were not antithetical [Huxley and Huxley, 1947]. Human morality, he went on to argue, was on an evolutionary track that led to a reshaping of the goals and ends of being human. The evolutionary destiny of human beings, he argued, is "to be the agent of the evolutionary process on this planet, the instrument for realizing new possibilities for its future" [Huxley, 1957].

The publication of E. O. Wilson's *Sociobiology* re-opened the flood gates and reinvigorated attempts to draw moral conclusions from the facts of evolution appeared. The naturalistic fallacy still loomed large as a roadblock in the minds of many. However, there has always been an underground of opposition to the alleged implications of the naturalistic fallacy. Some have argued that it is no fallacy and that values can be derived from facts under suitable derivational circumstances. Others have argued that although it is a fallacy, an evolutionary ethics need not commit it. Still others have challenged the very idea of the fact-value distinction that gives purchase to the 'fallacy'. Needless to say, the status of the fallacy and its implications for an evolutionary account of moral norms and behavior are controversial. It would be beyond the scope of this forum to sort all the ramifications of the current debate. Suffice it to say that there is no received consensus that has not been subject to severe critical examination.

Wilson's aim in *Sociobiology* was to create a new 'synthesis' that incorporated the social sciences under the wide umbrella of evolutionary biology. Reaction to this attempt was, to put it mildly, mixed [Segerstrale, 2001; Caplan, 1978]. While researchers were quite happy to explain non-human animal behavior in terms of

biological imperatives, it struck many as a stretch to extend the thesis to humans. On the one hand, this seems like a very non-Darwinian move to make since one of Darwin's major insights was to show the extent to which human beings were one with the other animals. On the other hand, the extrapolation of sociobiological insights to human activities faced some formidable challenges. For starters, the controlled experiments that were possible in principle with non-human animals were not available for the study of human beings. Thus, the contributions of genetics to human behavior could not easily be separated from the contributions of the environment. The sometimes bitter and acrimonious debate about the heritability of I.Q. and its significance for racial differences in humans is just one testament to this difficulty. Secondly, even if the genetic factors could be differentiated from the environmental factors, the fact is that a significant part of the environment of human beings is cultural rather than physical. We don't yet have a sufficiently rich understanding of human cultures to be able to identify and isolate what may or may not be the relevant variables that affect the acquisition and development of human behavioral repertoires.

So, where do we stand? From a Darwinian perspective, the capacity for the ability to recognize moral norms and to act on them is a phenotypic trait that was not present in our remote ancestors. It must, therefore, have evolved. Can we say more? Is it possible to 'derive' moral lessons or moral imperatives from the facts of evolutionary history? Here the message is mixed. Some say yes, others say no. When so many eminent scholars are in radical disagreement with one another we begin to suspect either that there are hidden agendas afoot or that there are several distinct questions that are being confounded. It is to the untangling of the distinct threads in this conversation that we now turn.

2.1 Two programs and two distinctions

EEM versus EET recast

Serious consideration of evolutionary epistemologies began with Donald Campbell's contribution to the Library of Living Philosopher's volume on Karl Popper [Campbell, 1974]. There were, to be sure, 19^{th} century predecessors but it was Campbell's seminal piece that made a strong case for taking an evolutionary approach to epistemology seriously. Both Campbell and Popper endorsed a view that saw the evolution of biological organisms and biological systems as on a continuum with the growth of human knowledge. I have argued elsewhere that this is a mistake. Instead, I proposed that there are two distinct but interrelated programs that seek to exploit biological and evolutionary considerations in an understanding of animal and human knowledge. I dubbed these two programs **EEM** (Evolution of Epistemic Mechanisms) and **EET** (Evolution of Epistemic Theses) [Bradie, 1986].

In *The Secret Chain*, I argued that there was a parallel between evolutionary ethics and evolutionary epistemology. I introduced a parallel distinction for programs in evolutionary ethics which I labeled **EMM** (for "Evolution of Moral

Mechanisms") and **EMT** (for "Evolution of Moral Theses"). Acronyms are useful until they get in the way of understanding. Since there is a parallel between evolutionary epistemology and evolutionary ethics and our present focus is on the emergence of *normativity*, it may be appropriate to introduce a new and simpler set of acronyms to cover both cases. On the one hand, there is the question of the evolution of the neural structures that enable us to form beliefs, generate knowledge, adopt norms and apply them. Call these programs or projects "**EM**" (evolution of mechanisms) projects whether they are concerned with the evolution of *epistemic* mechanisms or *moral* mechanisms. On the other hand, there is the question of the evolution of the norms themselves. The scientific method and the norms that guide it did not spring full blown from the head of Zeus. It has evolved over the centuries. Similarly there is a vast difference between the code of Hammurabi, the Golden Rule and the Universal Declaration of Human Rights. So there has been an evolution in our understanding of counts as moral norms as well. Concerns with the evolution of norms *per se* are part of "**EN**" (evolution of norms) programs or projects.

The **EM** program in epistemology is concerned with the use of evolutionary theory to explain the characteristics of the cognitive capacities of animals and humans. At some point in the distant past the Earth was populated by lineages consisting of organisms too primitive to be said to possess knowledge or have the capacity to reason. These lineages evolved over time into lineages that did consist of organisms with the capacity for reasoning and knowing. These capacities are phenotypic traits of the organisms that possess them and, given the general truth of the Darwinian picture, evolved under the impetus of natural selection and other evolutionary mechanisms. The details may be difficult to discover since brains and thoughts do not fossilize easily but anyone sympathetic to Darwinism will have little difficulty in accepting that this account of the evolution of the cognitive capacities of organisms is roughly correct.

The **EN** program in epistemology, on the other hand, is the attempt to account for the evolution of ideas, scientific theories and epistemic norms by using models and metaphors drawn from evolutionary theory. There are, no doubt, compelling analogies between the processes involved in biological evolution and the processes involved in the evolution of human knowledge. Karl Popper's "conjectures and refutations" model for the growth of scientific knowledge was consciously shaped by Popper's understanding of evolutionary theory [Hull, 1999]. Nevertheless, there are compelling disanalogies as well [Bradie, 1986]. The bottom line is that the case for epistemological **EN** programs is nowhere near as strong as that for epistemological **EM** programs. In fact, a little reflection shows that the fate of the two programs is independent of one another. There is every reason to suspect that some **EM** story will turn out to be correct. If it didn't, then this would necessitate a serious rethinking of Darwinian evolutionary theory. The same is not true for the **EN** programs. We don't as yet have a clear understanding of the development of human knowledge or the establishment of epistemic norms but it seems clear that no straightforward evolutionary story is going to provide it.

So much for the two dimensions of evolutionary epistemology. The moral of the story is that the same schema can be used to partition problems in the application of evolutionary theory to questions of ethics. The focus of **EM** research is on discovering and tracing the evolution of the neural mechanisms that enable human beings to formulate and act upon moral norms. The **EN** research traces the evolution of human moral systems as such and as exemplified in the work of Peter Kropotkin, Herbert Spencer, Julian Huxley, E. O. Wilson, Richard Alexander, Robert Hinde and others.

Again, we can be quite sure that some evolutionary account of the development of the neural mechanisms that underlie our capacity to be moral and to formulate norms is forthcoming, at least in principle. It would be nice if we could expect to find a fossilized brain structure that we could definitely associate with a proto-normative capacity but that is asking for what is probably impossible. But, this is a technical limitation and does not undermine our conviction that, from a broadly Darwinian perspective, *some* evolutionary account of the development of moral mechanisms is correct. The same cannot be said for providing an evolutionary account of the development of normative systems as such. Just as in the case of evolutionary epistemologies, I suspect that much of the warrant for accounts of the evolution of norms trades on our conviction that there is an evolutionary account of the emergence of the underlying mechanisms.

Phylogeny versus ontogeny

Complementing the distinction between **EM** problems and **EN** problems is a distinction that cuts across both. Biological organisms (or the lineages to which they belong) have an evolutionary history. Individual organisms have a developmental history as well. They mature from embryo to adult. Our biological understanding of organisms and their organs has both a *phylogenetic* and an *ontogenetic* dimension. Thus, consider the human brain. The structures that comprise the modern human brain are the phylogenetic descendents of structures in the brains and pre-brains of our evolutionary ancestors. The structures that comprise the modern *adult* human brain are the ontogenetic 'descendents' of the brain structure of the embryos that they developed from. The brain structures that enable cognitive and epistemic activity are no different from any other brain structures or any other organ for that matter.

Again a similar story can be told about the development of the neural structures underlying moral normativity. There is a phylogenetic story to be told as well as an ontogenetic story.

EN problems can be factored in the same way. Human knowledge has both a phylogenetic and an ontogenetic dimension. The growth of human knowledge over generations constitutes a phylogeny of sorts. Whether we are intrinsically more intelligent than our ancestral forebears — at least as far back as the development of art and language — is a matter of dispute. But there can be no argument over the fact that we know more than our ancestral forebears. Similarly, there can be

no doubt that adults know more than the embryos that they sprang from. The mechanisms contributing to the transmission of knowledge across generations and that are involved in the education of children obviously include cultural factors such as oral traditions, writing, records and training regimens.

A similar division applies to the development of moral norms. There is a historical account to be given of the development and intergenerational transmission of moral norms that constitutes a phylogeny of sorts. Complementing this account is an ontogenetic account of the emergence of moral awareness as the individual progresses from embryo to adult. Developmental psychologists have written much about these trends (e.g. [Kohlberg, 1984; Gilligan, 1982]).

Descriptive versus prescriptive accounts

Don Campbell saw the contributions of evolutionary theory to epistemology as constituting what he called a "descriptive epistemology" that was to *complement* but not replace the traditional "prescriptive accounts." [Campbell, 1988] He held that the empirical description offered by evolutionary theory of how we acquired the ability to acquire knowledge in conjunction with psychological accounts of how knowledge is acquired by individuals would, in combination with traditional accounts of the nature of justification, provide a complete picture of the nature of human knowledge.

Other naturalists took a more radical stand and proposed to replace the traditional accounts with purely empirical ones that eschewed the normative notion of justification altogether [Quine, 1969; 1995]. This was met by traditionally minded epistemologists with the charge, as formulated by Jaegwon Kim, that "for epistemology to go out of the business of justification is for it to go out of business" [Kim, 1988].

What is the relationship between evolutionary epistemology and traditional epistemology on the one hand and evolutionary ethics and traditional ethics on the other? I want to suggest that the cases are parallel and that in each instance there are three possible configurations for these relationships.

1. Evolutionary accounts are *competitors* to the traditional accounts. On this view, both are trying to address the same concerns and offering competing solutions to similar problems. Insofar as the tradition has been concerned with normative and prescriptive claims, the traditionalists have argued that evolutionary epistemology and evolutionary ethics are purely descriptive and thus fail to address these traditional questions of justification. Reidl [1984] is an epistemological example of this position. Herbert Spencer's *Principles of Ethics* and Julian Huxley's *Evolution and Ethics* are historical examples of this approach in ethics. More recently Robert Richards has proposed such an approach in *Darwin and the Emergence of Evolutionary Ideas on Mind and Behavior*.

2. Evolutionary epistemology and evolutionary ethics might be seen as *complementary* to the traditional approaches. On this view, the focus of traditional

epistemology and traditional ethics remains the justificational questions of the tradition. Evolutionary epistemology and evolutionary ethics supplement this account with an evolutionary account of the origin origins and development of norms. Campbell [1974] defends this view of epistemology. Westermarck's monumental *The Origin and Development of Moral Ideas* is the classic here for ethics. Darwin's treatment of morality in *The Descent of Man* falls roughly into this category.

3. Evolutionary epistemology and evolutionary ethics might be seen as *successor disciplines* to the traditions. On this reading, the evolutionary accounts do not address the traditional questions of traditional epistemology because it deems them irrelevant or unanswerable or uninteresting. Munz [1985] and Dewey [1910] are epistemological example of this position. Wilson [1978], Ruse [1986] and Ruse and Wilson [1986] defend this view for evolutionary ethics.

Obviously, the weight one attaches to evolutionary considerations in addressing the fundamental questions of normativity will depend on how one sees the relationship between the traditional accounts and the evolutionary accounts. I am not sure how one is supposed to defend one approach at the expense of another. For my own part, I think that a proper appreciation of our Darwinian heritage commits us to a hybrid version of option (1) and option (3). On the one hand, the evolutionary accounts do not merely contribute to our descriptive understanding of the emergence of norms but also attempt to address the prescriptive dimensions as well. On the other hand, the evolutionary perspective challenges some traditional notions of what counts as a 'justification' and thereby endorses a new vision of how we are to understand norms.

2.2 Substantive Considerations

Do evolutionary considerations tell us anything about the emergence of norms?

The evolution of moral norms.

In the *Descent of Man*, Darwin speculated on the origins of what he called our 'moral sense.' He argued that other intelligent organisms, were there any, would acquire a moral sense other than our own. Darwin cites the case of the hive-bees who might well support fratricide (*Descent*, chapter 4). According to Darwin, and I daresay, contemporary sociobiologists and evolutionary psychologists, the kind of organisms we are determines, in a broad sense, what kinds of norms we are likely to develop and endorse. This has both a positive and negative aspect.

On the positive side, given our social natures and the need for communal support in the raising of children, human beings have evolved altruistic motivations that temper inclinations toward self interest. We can well imagine, in the spirit of

Darwin, that other creatures that are intelligent but self reliant would not be moved by considerations of sympathy and empathy with their fellow kind.

On the negative side, just as the Naturalistic Fallacy suggests that 'is' does not imply 'ought,' it is often pointed out in ethical circles that 'ought' implies 'can.' The idea is that no norms that require what is impossible can be binding on us. So, for instance, it is folly to establish or endorse norms that are beyond our capacity to obey. The norm "Thou shalt not kill" seems perfectly proper while the norm "Thou shalt not eat" seems ludicrous. There are limitations to the expectations we can have for ourselves and for others. These limitations are a result of our limited physical, emotional and intellectual capacities. But these limits are the fruits of our evolutionary progress. So, it seems reasonable that the evolved limitations of our physical and mental capabilities ar relevant to determining or setting the boundaries of our normative demands.

Some moral theorists might take exception to the above conclusion. A God-centered ethics might argue that the limitations of human beings are the reflection of original sin or something of the sort and that this just shows that human beings need to resign themselves to the will of their Maker. A secularized version of such an ethic can be found in Kant who postulates an ideal Kingdom of Ends as the (ultimately) unachievable model for human moral behavior. These concerns can not be easily dismissed although I do not propose to pursue them here. Instead I commend to your attention James Rachels' *Created from Animals* which explores the implications of Darwinism for formulating a moral theory and effectively calls into question both theologically based and Kantian ethical positions.

Rachels' book is one long argument to the effect that Darwinism undermines the concept of human dignity that he claims forms the basis for traditional moralities. This, in turn, has implications for the moral status of animals. Rachels takes what he calls the traditional concept of human dignity to be the presumption that the primary purpose of morality is the "protection of human beings and their rights and interests" [Rachels, 1990]. This presumption is supported by certain factual (or quasi-factual) assumptions about human nature. Two basic claims emerge from this factual base and support the sanctity of human dignity. One is the presumption that human beings were created (as special) in the image of God. Rachels calls this the "image of God thesis." The second is the presumption that human beings alone among the animals are rational beings. It does not follow logically from these presumptions that human dignity is or ought to be the lynch pin of morality. But, Rachels argues, the primacy of human dignity does rest on and is supported by these presumptions. They serve, as it were, as the rationale for putting human concerns ahead of all others in matters of morals.

Darwinism indirectly undermines the primacy of human dignity by undermining the presumptions that support the doctrine. The Darwinian perspective marginalizes God as the creator of human beings as special. Although Darwinism does not entail that God did not create human beings as special, it renders the story superfluous or suspect.[1] From the Darwinian perspective, humans are just one among

[1]If one does think the theory of evolution entails that God did not create human beings as

the animals. The Darwinian theory of common descent suggests that all organisms are interrelated. Darwinian gradualism suggests that differences between species are often matters of degree and not matters of kind. These implications undermine the status of human beings as special and in so doing undermine the traditional moralities which are based on that explicit or implicit assumption.

To replace the discarded image with something of value, Rachels proposes a view he calls "moral individualism." Moral individualism treats all individuals, human or not, as individuals and not as members of a certain species. Considerations of moral relevance are to be determined by circumstances and not by fiat. Rather than pursue that development here I want to note that Rachels' argument is not intended merely to replace one set of moral norms by others but that it calls into question some of the fundamental assumptions that lie behind any norms. This takes us into the realm of the meta-ethical.

What, if anything, are the implications of Darwinism for meta-ethics? The verdict is still out but one can find adherents of a wide diversity of views. Michael Ruse, for one, has argued that a Darwinian approach to ethics rules out any form of moral realism in favour of an "error" theory of the form first promoted by David Hume [Ruse, 1986]. This has led to a vigorous debate in the literature with no clear resolution in sight.

One might argue that Darwinism lends itself to moral realism by adapting an argument formulated by Donald Campbell with respect to human cognitive faculties. Campbell argues that just as the physical environment shapes the evolution of organismic features, so the physical environment is held to shape the evolution of what we know [Campbell, 1974]. Our cognitive faculties and our scientific theories 'fit' our physical environments in much the same way the organisms in successful lineages are co-adapted to their environments. If our cognitive abilities and guesses about the world we live in were not on the mark more often than not we would be on the road to extinction. There is a congenial reciprocity between what we think and how we think and what we think about. Campbell calls this view 'critical realism' and thinks that a Darwinian viewpoint is committed to it. I have some reservations about this argument in its guise as an account of the evolution of our cognitive capacities but, were one persuaded by it, it might be invoked as a defense of the contention that Darwinism is committed to or, at least, is compatible with some form of moral realism.

If we understand a minimal version of moral realism to be committed to the view that there are 'moral facts' in the world (whatever that might mean) then we can well imagine that the 'moral environment' might shape the evolution of our moral capacities and moral norms in much the same manner as the physical environment is held to shape our cognitive capacities and cognitive norms. I'm not sure how far this argument can be pushed but it seems that the cognitive realm and the moral realm are, *prima facie*, on a par and if an evolutionary argument for critical *physical* realism can be made then perhaps an evolutionary argument

special, one need only reflect on the possibility that the entire evolutionary process is just God's way of insuring the specialness of human beings through "secondary causes."

for critical *moral* realism could be made as well.

This ignores, of course, all the arguments that have been made to the effect that moral claims have an absolutely different status from physical claims and I am far from suggesting that an appeal to evolutionary theory is likely to resolve this debate. In fact, since I do not think that Campbell's argument should persuade us of the truth of "critical physical realism" as he understands it, I do not think a parallel argument would persuade anyone of the truth of "critical moral realism" either. With respect to the ultimate status of moral claims evolutionary theory is, to this point, silent.

The evolution of cognitive norms

Just as the latter part of the nineteenth century saw a cottage industry of attempts to apply Darwinian insights to matters political, social and moral, so too the beginnings of what Donald Campbell was to call 'evolutionary epistemology' can be found in the latter part of the nineteenth century as well. I have spilled a lot of ink on the history of the development of evolutionary epistemology over the years and rather than rehearse it all here again I commend to your attention the excellent surveys by Campbell [1974] and myself [Bradie, 1986].

Here I just want to suggest that the Darwinian impact on our understanding of the evolution of cognitive norms has, in parallel with the Darwinian impact on our understanding of moral norms, both a positive and a negative dimension. On the positive aspect, we can cite the work of Nick Rescher and Michael Ruse. Rescher distinguishes between what he calls "thesis Darwinism" and "methodological Darwinism" [Rescher, 1990]. He endorses the latter but not the former. Ruse goes so far as to suggest that the adoption by nineteenth century philosophers of science of the norms of the modern scientific method have a Darwinian explanation. In the course of developing this thesis he tells an amusing parable about two ancestors looking for a cave to spend the night and encountering a band of sabre toothed tigers entering a likely cave. All but one of the tigers eventually comes out. One of the pair, with limited mathematical skills, enters the cave. The other demurs. Ruse asks: which one do you think was your ancestor?

I have reservations about the cogency of all such stories as I think they conflate the two programs that I have here distinguished as **EM** and **EN**. Evolutionary stories about the development of the physical and mental mechanisms that enable us to count and to reason are one thing but evolutionary accounts of the emergence of specific norms or theses are quite another. There is a strong analogical pull to co-opt what evidence there is for the former in support of the latter but it is not clear to me that doing so throws any explanatory light on the emergence of theses or norms.

On the negative side, I want to go back and draw some implications from Dewey's 1910 article. In that piece, Dewey argued that Darwinism properly understood undermined many of the traditional pursuits of philosophers. He thought, in particular, that the search for essences was a lost cause in the light of the evolution-

ary history of human kind. He went on to develop an 'instrumental' conception of knowledge that saw concepts and theories as 'tools' that we develop to cope with our environments. I do not know whether he explicitly invoked Darwin in support of this but the view certainly has a Darwinian flavour that Don Campbell, for one, might have approved of. In any case, Dewey went on to challenge traditional epistemological thinking in a number of books including his pragmatic challenge to traditional epistemology, *The Quest for Certainty* [Dewey, 1960]. In that book, he challenged the correspondence theory of truth and knowledge and promoted his own instrumental views. The quest for certainty, most familiar to moderns as a legacy of the Cartesian tradition, is a version of epistemological essentialism — a search for the unshakable foundations of human knowledge. It creates impossible demands for knowers to satisfy and in this respect parallels the impossible demands of traditional moral theories. In an age where we could believe that we had a touch of the divine, one could imagine that epistemic certainty might be realizable if only one was prepared to follow the proper regimen. But once the epistemic 'image of God' thesis had been undermined by Darwinian theory, the quest for certainty appears to be chimerical. What to put in its place? Nineteenth century pragmatists promoted, although they did not invent, the epistemic doctrine of fallibilism that can be understood as the doctrine that all foundations are temporary respites. This view seems quite compatible with a Darwinian perspective on the world. Evolved human beings have finite limitations and are subject to errors and mistakes. We live in a constantly evolving world and need to re-adapt to changing circumstances as they occur. Under such conditions, the quest for certainty appears as folly. Darwinism does not *entail* that it is a mistake but it does seem to undermine the grounds for taking the quest seriously in much the same way that Rachels argued that Darwinism undermines traditional absolutist moral theories.

Can evolutionary considerations justify norms?

Some naturalists such as the American philosopher W. V. O. Quine, at least in some of his writings, suggested that epistemology (and, one presumes, ethics as well) should abandon the demand for justification in favor of giving a bio-psychological account of how we acquire knowledge. This is consistent with Donald Campbell's call for a 'descriptive epistemology' that would provide a scientific account of the evolution of our cognitive faculties and of our methods of knowledge acquisition. Campbell, of course, thought that a descriptive epistemology rooted in evolutionary biology and developmental psychology would complement the normative approach of traditional philosophers. Quine, when pressed, provided an 'instrumental' view of normativity that drew on the analogy between engineering and science. Some Popperians, such as Peter Munz, have argued that the fallibilistic implications of Darwinism undermine the rationale for any attempt to provide justifications [Munz, 1993]. And, of course, Popper is well known for arguing that we are never 'justified' in believing anything – the best we can hope

for is to produce fruitful hypotheses that withstand rigorous tests.

If we are to take Dewey's insights seriously, then we have to question the viability of the notion of justification in a post-Darwinian age. A persuasive argument can be given for abandoning any 'classical' conception of justification as resting on *a priori* foundations or any essentialistic conception of knowledge or morality. Following Rachels, Darwinism does not entail that such pursuits are hopeless but it does serve to 'undermine' the metaphysical picture that grounds those pursuits. It does not follow, however, that the concept of justification has be abandoned *tout court*. The rise of modern science saw the decline of the concept of knowledge as demanding certainty but not the abandonment of 'justifications.' The absolutist conception of knowledge was replaced by a fallibilistic conception. Epistemic fallibilism does not eschew justification *per se*; it just rejects the idea that any justifications are absolute. As, Wilfrid Sellars once put it (more or less), knowledge has foundations but those foundations are at best provisional and subject to reassessment as our understanding of what we are investigating matures. We are, as it were, always grounded but always prepared to challenge those grounds should the need to do so arise [Sellars, 1997].

A similar case can be made for a 'normative fallibilism.' The rejection of absolutist moralities does not mean the rejection of all morality. It just means that we adopt norms consonant with our agreed upon ends but stand ready to modify and challenge those ends as the need might arise. This view, I take it, is consonant with Dewey's ethics and Quine's 'engineering' approach.

Of course, once one rejects 'ultimate' ends and 'essential' goods, one is subject to the charge that abandoning the absolute means endorsing relativism. But this charge is spurious. Relativism is rampant when *all* views are equally legitimate at *all* times. Fallibilism does not countenance such a relativistic slant. At any given stage of our reasoning, be it epistemic or moral, there are *mutually accepted grounds* that circumscribe what it is reasonable to believe or act on. The difference between the fallibilist and the absolutist is that the fallibilist is prepared, under appropriate circumstances, to challenge the adequacy of the mutually accepted grounds and possibly to move on to other vantage points. Of course, this response invites a challenge to specify what is to count as 'appropriate' and 'adequate' but no absolute reply is forthcoming. These are matters for continual discussion and (re-)evaluation. For a fallibilist that is the best one can hope for.

Does an evolutionary perspective *entail* that we adopt such a view? No, the best one can say for the future of our discussions of the nature of normativity is that the evolutionary point of view lays the groundwork for a metaphysical picture that *supports* a thoroughgoing epistemological and moral fallibilism.

3 FINAL THOUGHTS

The relevance of evolutionary theory and the facts of evolutionary history to understanding the nature of normativity turn on the answers to two preliminary questions.

First, are we to follow Wittgenstein and eschew all empirical input for the resolution of philosophical problems or are we to follow Dewey in thinking that the advent of Darwinism radically reshapes the philosophical landscape and forces us to rethink some of the timeworn and traditional problems that philosophers have posed for themselves? I have suggested that Dewey's path is the one we ought to follow but I freely admit that there is no knock down drag out argument that will convince the diehard traditionalist. Second, we must take a stance on the relationship between the traditional accounts and the evolutionary accounts. I have advocated a hybrid view that recognizes the importance of normative demands but sees evolutionary theory as forcing a reconceptualization of the nature of those demands.

That said there are a number of open questions still left to be pursued.

What are the biological bases of normative activity? What are the neural processes that enable us to be the normative creatures that we are? How have we come to accept the norms that we do? This is a historical and cultural history project more than anything else.

Beginning in the latter part of the nineteenth century and continuing to the present day, there have been numerous attempts to apply Darwinian ideas to the evolution of morality. Needless to say many of these attempts were as much a product of the author's views about how things should have turned out as they were accurate accounts of how norms were adopted and discarded over the centuries. I am far from supposing that a completely unbiased historical account of the development of norms is possible. Nevertheless, it seems to me to be a continuing concern that needs to be periodically addressed is our understanding both of normativity and of the historical contingencies that shape the evolution of norms changes. As a first but daunting step, it would be very interesting and constructive if someone or some group were to undertake the task of constructing a meta-analysis of the analyses that have already been written. There is more than enough work to keep future generations of scholars busy.

BIBLIOGRAPHY

[Alexander, 1987] R. Alexander. *The Biology of Moral Systems*. Aldine De Gruyter: New York, 1987.

[Bagehot, 1956] W. Bagehot. *Physics and Politics; or, Thoughts on the Application of the Principles of "Natural Selection" and "Inheritance" to Political Society*. Boston: Beacon Press, 1956.

[Bradie, 1986] M. Bradie. Assessing Evolutionary Epistemology. *Biology & Philosophy* 4: 401-59, 1986.

[Bradie, 2001] M. Bradie. The Metaphysical Foundation of Campbell's Selectionist Epistemology. In *Selection Theory and Social Construction: The Evolutionary Naturalistic Epistemology of Donald T. Campbell*, edited by David Hull and Celia Heyes. Albany: State University of New York Press, 2001.

[Bradie, 1990] M. Bradie. Should Epistemologists Take Darwin Seriously? In *Evolution, Cognition and Realism: Studies in Evolutionary Epistemology*, edited by N. Rescher, 33 - 38. Lanham, MD: University Press of America, 1990.

[Bradie, 1994a] M. Bradie. Epistemology from an Evolutionary Point of View. In *Conceptual Issues in Evolutionary Biology*, edited by Elliott Sober, 453-75. Cambridge, MA: The MIT Press, 1994.

[Bradie, 1994b] M. Bradie. *The Secret Chain: Evolution and Ethics*. Albany, NY: State University of New York Press, 1994.

[Campbell, 1988] D. T. Campbell. Descriptive Epistemology: Psychological, Sociological, and Evolutionary. William James Lectures, Harvard University. In *D. T. Campbell, Methodology and Epistemology for Social Science: Selected Papers*, edited by E. S. Overman, 435-86. Chicago: University of Chicago Press, 1988.

[Campbell, 1974] D. T. Campbell. Evolutionary Epistemology. In *The Philosophy of Karl R. Popper*, edited by P. A. Schilpp, 412-63. LaSalle, IL: Open Court, 1974.

[Campbell, 1991] D. T. Campbell. A Naturalistic Theory of Archaic Moral Orders. *Zygon* 26, 1: 91-114, 1991.

[Caplan, 1978] A. Caplan, ed. *The Sociobiology Debate*. Harper & Row: New York, 1978.

[Changeaux, 1985] J.-P. Changeaux. *Neuronal Man*. New York: Pantheon, 1985.

[Darwin, 1981] C. Darwin. *The Descent of Man, and Selection in Relation to Sex*. Princeton University Press, New Jersey, 1981.

[Dewey, 1910] J. Dewey. *The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought*. New York: Henry Holt & Co., 1910.

[Dewey, 1960] J. Dewey. *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. New York: Capricorn Books, 1960.

[Edelman, 1985] G. M. Edelman. *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books, 1985.

[Gilligan, 1982] C. Gilligan. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press, 1982.

[Hinde, 2002] R. A. Hinde. *Why Good Is Good: The Sources of Morality*. London; New York: Routledge, 2002.

[Hull, 1999] D. Hull. The Use and Abuse of Karl Popper. *Biology and Philosophy* 14: 481-504, 1999.

[Huxley, 1957] J. Huxley. *Knowledge, Morality and Destiny*. The New American Library: New York, 1957.

[Huxley, 1989] T. H. Huxley. Evolution and Ethics. In *Evolution and Ethics*, edited by James Paradis and George Williams. Princeton, New Jersey: Princeton University Press, 1989.

[Huxley and Huxley, 1947] T. H. Huxley and J. Huxley. *Touchstone for Ethics: 1893-1943*. New York: Harpers, 1947.

[Kim, 1988] J. Kim. What Is 'Naturalized Epistemology'? *Philosophical Perspectives 2. Epistemology*: 381-405. 1988.

[Kohlberg, 1984] L. Kohlberg. *The Psychology of Moral Development: The Nature and Validity of Moral Stages*. 1st ed. San Francisco: Harper & Row, 1984.

[Kropotkin, 1919] P. Kropotkin. *Mutual Aid: A Factor of Evolution*. New York: Alfred A. Knopf, 1919.

[Kropotkin, 1924] P. Kropotkin. *Ethics: Origin and Development*. Translated by Louis S. Friedland and Joseph R. Piroshnikoff. First English ed. New York: The Dial Press, 1924.

[Mackintosh, 1834] Sir James Mackintosh. *A General View of the Progress of Ethical Philosophy*. Philadelphia, PA: Carey, Lea and Blanchard, 1834.

[Manier, 1978] E. Manier. *The Young Darwin and His Cultural Circle*. D. Reidel Publishing Company: Dordrecht, Holland, 1978.

[Moore, 1962] G. E. Moore. *Principia Ethica*. Cambridge University Press: London, 1962.

[Munz, 1985] P. Munz. *Our Knowledge of the Growth of Knowledge: Popper or Wittgenstein*. London: Routledge & Kegan Paul, 1985.

[Munz, 1993] P. Munz. *Philosophical Darwinism: On the Origin of Knowledge by Means of Natural Selection*. New York: Routledge, 1993.

[O'Hear, 1997] A. O'Hear. *Beyond Evolution: Human Nature and the Limits of Evolutionary Explanation*. Oxford: Clarendon Press, 1997.

[Popper, 1968a] K. R. Popper. *The Logic of Scientific Discovery*. New York: Harper, 1968.

[Popper, 1968b] K. R. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper, 1968.

[Popper, 1976] K. R. Popper. Darwinism as a Metaphysical Research Programme. *Methodology and Science*, 9:103-19, 1976.

[Popper, 1978] K. R. Popper. Natural Selection and the Emergence of Mind. *Dialectica* 32: 339-55, 1978.
[Popper, 1984] K. R. Popper. Evolutionary Epistemology. In *Evolutionary Theory: Paths into the Future*, edited by J. W. Pollard, 239-55. London: John Wiley & Sons Ltd., 1984.
[Quine, 1969] W. V. Quine. Epistemology Naturalized. In *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1969.
[Quine, 1995] W. V. Quine. Naturalism; or, Living within One's Means. *Dialectica* 49, 2-4:251-61, 1995.
[Rachels, 1990] J. Rachels. *Created from Animals: the Moral Implications of Darwinism*. Oxford: Oxford University Press, 1990.
[Radnitzky and Bartley, 1987] G. Radnitzky and W. W. Bartley. *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. La Salle, Illinois: Open Court, 1987.
[Rescher, 1977] N. Rescher. *Methodological Pragmatism*. Oxford: Basil Blackwell, 1977.
[Rescher, 1990] N. Rescher. *A Useful Inheritance: Evolutionary Aspects of the Theory of Knowledge*. Savage, MD: Rowman & Littlefield, 1990.
[Richards, 1987] R. J. Richards. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press, 1987.
[Riedl, 1984] R. Riedl. *Biology of Knowledge: The Evolutionary Basis of Reason*. John Wiley & Sons: New York, 1984.
[Rorty, 1979] R. Rorty. The Unnaturalness of Epistemology. In *Body, Mind and Method*, edited by D. F. Gustafson and B. Tapscott, 77-92. Dordrecht: D. Reidel, 1979.
[Ruse, 1986] M. Ruse. *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*. New York: Basil Blackwell, Inc., 1986.
[Ruse and Wilson, 1986] M. Ruse and E. O. Wilson. Moral Philosophy as Applied Science. *Philosophy* 61: 173-92, 1986.
[Segerstråle, 2001] U. Segerstråle. *Defenders of the Truth: The Sociobiology Debate*. Oxford: Oxford University Press, 2001.
[Sellars *et al.*, 1997] W. Sellars, R. Rorty, and R. Brandom. *Empiricism and the Philosophy of Mind*. Cambridge, Mass.: Harvard University Press, 1997.
[Sidgwick, 1964] H. Sidgwick. *Outlines of the History of Ethics for English Readers*. Beacon Press: Boston, 1964.
[Spencer, 1966] H. Spencer. *The Principles of Ethics*. Reprint of the 1879 edition ed. Vol. I. Osnabrück: Otto Zeller, 1966.
[Westermarck, 1906–8] E. Westermarck. *The Origin and Development of the Moral Ideas*. New York: The Macmillan Company., 1906-8.
[Wilson, 1975] E. O. Wilson. *Sociobiology*. Cambridge, MA: Harvard University Press, 1975.

This page intentionally left blank

EVOLUTIONARY ETHICS

Catherine Wilson

1 INTRODUCTION

Evolutionary Ethics is the study of the relationship between the theory of natural selection and ethical theory and practice. It is directed to topics such as the following: (1) The implications of Darwinian evolution for the epistemological status of moral claims. (2) The significance of our evolutionary history for an understanding of moral practices and institutions. (3) The relevance of information about human evolution to the formation of morally justifiable social policies and to individual decision-making.

Moral philosophers have tended on the whole to skepticism on these issues. Moral theory, they may observe, is a product of the enlarged human neocortex, and the psychological and behavioural innovations associated with encephalisation include rationality, language, impulse control, objective knowledge of causes and effects, and the ability to represent internally, both verbally and pictorially, absent, past, and nonactual persons and events. "Rape" in ducks and "adultery" in bluebirds, however interesting, has little in common with the human proclivities bearing the same names. Further, variations in individual personalities and the great diversity of cultural practices seem to throw into question any attempt to establish an underlying human nature and to decide what constitutes a distortion or a remodeling. In place of a novel ethical code reflecting a dispassionate state of scientific enlightenment about human needs, evolutionists seem in their normative moments to parade values — or at least to sigh over inevitabilities — that have an authoritarian and archaic case to them: e.g., individualism, hierarchy, male dominance, the double standard in sexual morality, nationalism, xenophobia and war [Singer, 1982].

Many social species, it will be observed, interact with one another in ways that remind observers of either good moral behaviour or infringements of decency. Their members possess the capacity to form alliances, to categorize others as mates, friends, enemies, leaders and subordinates, and they may share food, tend one another's wounds, co-operate in hunting, nest-building, and feeding the young, and put themselves at risk to nourish and defend their offspring, as well as attacking, deceiving, and displacing one another. They also may also punish offenders, or engage in acts of relationship repair, mediating fights or expressing contrition [Smuts, 1985; De Waal, 1996; Flack and De Waal, 2000]. Yet, animal behaviour, it is sometimes objected, cannot be termed moral, since it does not reflect concepts

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

of the self as an agent capable of performing noble and selfish actions, and since nonhuman social animals do not engage in moral education, or seek moral progress. Only human beings construct representations of themselves as actors in the world and posit such abstract entities as rights, responsibilities, and legitimate interests that nonhuman animals cannot possess. Human beings are the only creatures to advance moral theories and to enter into contests and competitions regarding their truth, and their prehuman evolutionary history is thought to shed little or no light on these intellectual and political activities.

The kinship of man and animal, the evolution of higher forms from lower, and the relation of morality to sentiment and instinct was first discussed in the 18th century, with Immanuel Kant taking a decidedly critical position with respect both to evolutionary thought and to what he considered the impure influence of natural history, anthropology, and psychology on ethics. The lingering influence of Kant's critique of naturalism and his insistence that moral theory requires reference to supersensible concepts such as the noumenal will has been marked down to the present day. None of arguments above seems, however, decisive against the very possibility of Evolutionary Ethics. While no claim about the behaviour of members of other species permits any inference regarding our own, the usefulness of a comparison class depends upon the terms of the question being addressed. Neither individuality nor the differences between groups implies the impossibility of a study of human nature, understood as a study of the characteristics that we share with other species and the characteristics that are differentiating. Ideological extrapolations, while common enough to be worrisome, do not preclude the possibility of nonideological extrapolations of the same or different data. And while moral theory can be regarded as necessarily having reference only to human actions, this does not imply that general information that is not bound by the human context is of no use in articulating moral theory. The study of proto-morality, as it might be called, in non-linguistic creatures without conscious moral concerns reveals the platform for morality, the psychological receptivities and dispositions without which it is impossible to imagine the formation of explicit moral codes.

Towards the middle of the 19th century, Darwin came to believe that the theory of natural selection had reconciled the moral sense of the ethical intuitionists of his day with utilitarianism [Barrett, 1987, 609]. Darwin speculated that "any animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well, or nearly as well developed, as in man" [Darwin, 1990, 83]. He regarded human character traits as shaped by agonistic interaction between tribes, with the moral virtues of courage, sympathy and faithfulness offering a competitive advantage [ibid., 110]. He did not however consider tribal warfare a permanent condition of the species and he regarded slavery with repugnance. Darwin was impressed by "numerous points of mental similarity" and "the close similarity between the men of all races in tastes, dispositions, and habits" [Darwin, 1990, 153f.] He was nevertheless conflicted about the over-protection of the weak, and fell back on the view that they tended to

self-eliminate. He seemed at once to recognize and fear an objective condition of degeneracy and to deny that evolution had any particular goals. [Crook, 1994, Ch. 1]. The notion that the "survival of the fittest" was a bio-historical law underwrote metaphysical systems such as that of Herbert Spencer. The partitioning of Africa amongst European colonial powers, and the extermination of all but a small fraction of the American Indian population begun in the colonial era and completed by the U.S. army in the late 19th century, were often viewed as exemplifying a natural process of selection, accelerated by human effort and leading to overall improvement. Whether modern man was to be considered wild or tame, whether he degenerated under conditions of domestication and peace, and the baleful effect of the "herd instinct", as Nietzsche termed it, were much-debated questions towards the end of the 19th century. These questions have little current resonance.

The notion that capitalism and militarism were biological necessities or represented higher points of development was contested virtually from Darwin's day onwards [Crook, 1994, 153ff.]. Thomas Henry Huxley, citing Buddhist and Stoic traditions of ethics, agreed with Darwin regarding the existence of an evolutionary basis for morality, but declared in his Romanes Lectures that the struggle to exist at the expense of others was frankly opposed to morality: "[T]he practice of that which is ethically best — what we call goodness or virtue — involves a course of conduct which, in all respects, is opposed to that which leads to success in the cosmic struggle for existence. In place of ruthless self-assertion it demands self-restraint; in place of thrusting aside, or treading down, all competitors, it requires that the individual shall not merely respect, but shall help his fellows; its influence is directed, not so much to the survival of the fittest, as to the fitting of as many as possible to survive. It repudiates the gladiatorial theory of existence." [Huxley, 1893, p. 33] The loss of the flower of youth in the First World War threw the doctrine of the survival of the fittest into confusion, and triumphalist interpretations of natural selection ceded over the course of the century to universalist and egalitarian theories but also to existentialism and nihilism. Awareness of the horrifying results of the translation of the American Eugenics movement to Fascist Germany [Nicholas, 2005] brought notions of manifest racial destiny and genetic improvement into permanent disrepute following World War II. C.H. Waddington, the eminent developmental biologist, nevertheless surprised readers of *Nature* with an article published in 1941 stating that science was in a position to provide mankind with a true system of ethics. Waddington referred to "the direction of the evolutionary process as a whole", and maintained that helps in the right direction were by definition morally good and distractions from this "evolutionary course" were evil. The notion that the process of evolution incorporated a bias towards the production of superior forms remained in the popular mind as well. It was dealt a significant blow in the wider culture by Jacques Monod's *Le Hasard et la Necessité* [1971], a bestseller that was immediately translated into English. "Man knows at last that he is alone in the universe's unfeeling immensity out of which he emerged only by chance."

Though future intellectual historians may perceive a connection between the

allegedly failed ideals of the benevolent, expensive, and politically unstable Welfare State and the return to nature, viewed as an arena of untrammeled competition, and while traces of Social Darwinism remain in the form of the inference that the dominant owe their position to their inherited superior characteristics, rather than chance and history, contemporary Evolutionary Ethics has adopted more authentically Darwinian views and questions. The emphasis in the literature has shifted from the issues of war and ethnic conflict to social interaction in pairs and groups. Public attention has been captured by an array of semipopular books on animal ethology, of an openly prescriptive or at least highly suggestive character.

"The new sciences of human nature", Steven Pinker ventures, "can help lead the way to a realistic, biologically informed humanism ....They promise a naturalness in human relationships, encouraging us to treat people in terms of how they do feel rather than how some theory says they ought to feel. They offer a touchstone by which we can identify suffering and oppression wherever they occur ... They renew our appreciation for the achievements of democracy and of the rule of law. And they enhance the insights of artists and philosophers who have reflected on the human condition for millennia." [Pinker, 2002, xi] Some philosophers have followed this lead. Daniel Dennett asks in *Darwin's Dangerous Idea*, "From what can 'ought' be derived?" " The most compelling answer", he decides, "is this: ethics must somehow be based on an appreciation of human nature — on a sense of what human nature is or might be like and what a human being might want to have or be" [Dennett, 1996, 268]. As Mary Midgely points out, all prescriptive moral doctrines are based on some theory of what people are like, and the claim that we have no nature really amounts to the claim that we are "— naturally — quite plastic" [Midgely, 1978, 166].

Yet the supposition that the theory of evolution can provide not only unique and valuable insights into human nature but offer guidance in framing appropriate norms of conduct remains problematic. Though a large literature has developed around applications of game theory to the study of the evolution of social behaviour, no professional philosopher has written a standard work in Evolutionary Ethics providing a rigorous, yet favourable treatment of the subject. This is not only because, as is sometimes alleged, philosophers since Descartes have promoted a misleading view of human beings as dematerialized reasoning, willing, choosing entities, who control and alter nature without being subject to it, but because conceptual difficulties above and beyond those just set aside infest the very idea of an Evolutionary Ethics. To appreciate the potential value of evolutionary approaches and to understand their limitations, it will be useful to sketch the central assumptions behind the positive assessments just quoted, as well as indicating the importance of some underutilized conceptual resources from recent evolutionary theory, including the notions of the extended phenotype, obsolescence, and parasitism.

2 BACKGROUND ASSUMPTIONS

The central assumption of evolutionary ethics is that genes direct the construction of the body and the brain, producing, maintaining, and eventually shutting down, according to timetables for development and senescence, morphological structures, physiological mechanisms, and behavioural features in living organisms and that those genes that produce forms, processes and structures more conducive to their own survival and replication than other genes tend to dominate in populations. This insight can be expressed in terms of competition and rivalry amongst alleles of "selfish" genes [Dawkins, 1976, 2]. An individual organism is on this view a machine endowed with certain behavioural propensities and competencies that a federation of genes has jointly constructed, each acting in its own interest, however much it is constrained or assisted by its neighbours. Summarizing:

1. Genes contain the instructions for building organisms that have morphology, exhibit behaviour, and that construct species-typical artifacts.

2. Morphology and behaviour are shaped by selection for particular genes, some of which are more effective replicators than their variants.

3. Overt competition for food, water, shelter, reproductive advantage, etc. amongst members of a breeding community or between groups plays some role in evolutionary change.

4. Sexual reproduction in a species implies the existence of two distinct genotypes (and variants of them) and varying degrees of sexual dimorphism in the phenotypes. In humans, the evolutionary process has resulted in some degree of dimorphism, not only in bodily structure and physiology, but in cerebral organization, hence in dispositions and capabilities.

5. The number of females in breeding community has more influence on population dynamics than the number of males; the fertility of females constitutes a "limiting resource" for males.

6. The human mode of life incorporates features such as food-sharing, mutual assistance, tribal loyalty, partiality to kin, social dominance of some humans by others, protracted care of the young by mothers and other females, and some degree of paternal care.

From these theses, it is argued, one may explain certain features commonly observed in human social systems that might surprise a rational but scientifically uninformed visitor from another planet. Such features might include the division of labour, hierarchical and political systems, sibling rivalry, parental repression alternating with parental devotion, sexual jealousy, monogamy punctuated by infidelity, and hostility to strangers. Humans who possessed too little by way of the dispositions and emotions that sustain these practices failed to pass on, it is argued, as many genes as those who possessed them.

Before going on to consider the possible bearing of such evolutionary explanations of the origins of certain traits on moral prescriptions, some points relating to the explanatory claims of evolutionary theory should be clarified.

First, the doctrine of the selfish gene does not imply that persons are selfish and are determined to be so by their genes. "Vernacular" competition is salient in some species and in some contexts, but, according to 3. above, it is not to be confused with the purely metaphorical strivings of genes to be present in future generations. An organism's genetically influenced selfishness is a function of the contribution of selfish behaviour under the control of any gene to its own replication. If "selfish" genes for selfish behaviour thrive in a given environment, organisms will be selfish; if "selfish" genes for unselfish behaviour thrive in a given social environment, then individual organisms will be unselfish. Second, while, to the layman, the shell of a snail is a morphological feature produced by its genes, and different from the nest of the bird, Dawkins's concept of the extended phenotype [Dawkins, 1982] implies that genes can act at a distance from the organism in which they reside, exerting their effects over the environment and even over other organisms. Potentially, numerous constructions of organisms that are directed by a gene and favourable to its replication count as phenotypic components of *some* organism. Insofar as much of human life is dedicated to the making of a physical and cultural environment, the concept of the extended phenotype may be consequential and deserves further philosophical exploration. While it is possible and important in some contexts to make a distinction between innate and learned features, or natural predispositions and cultural acquisitions, the boundaries between these concepts are not sharp: the ability to learn, certain dispositions to learn, and what can be learned are innately determined, and the formation of cultures with particular typical characteristics is, for humans, natural. Third, not all characteristics of the phenotype are adaptive. This is so not only because some may be accidental side effects of genuinely adaptive traits, but on account of the ubiquity of parasitism, a phenomenon which, as Dawkins points out, implies the falsity of the claim that "it is useful to expect individual organisms to behave in such a way as to maximize their own inclusive fitness, or in other words to maximize the survival of copies of the genes inside them." [ibid., 55] Certain morphological, physiological, and behavioural characteristics that appear typical in a species need not aid them in their competition to survive and reproduce. Instead, they may be elicited by parasites possessed of heritable powers to invade and influence. Dawkins suggests that "animals exert strong power over other animals, and ... frequently an animal's actions are most usefully interpreted as working in the interests of another animal's inclusive fitness, rather than its own" [ibid., 68]. The cuckoo, for example, has discovered how to look and act in such a way as to produce an "addiction" to feeding it in its helpless foster parents. Accordingly, while stereotypical behaviour in a species, or in some of its members, may signal that it is a product of evolution, we are not entitled to assume that it favours the survival and reproduction of the animal in which it is observed. A fourth concept of importance is obsolescence. "The animal we are looking at", says Dawkins, "is probably out of

date, built under the influence of genes that were selected in some earlier era when conditions were different" [ibid., 35]. Questions about the adaptive significance of some currently-observed form of behaviour — based on the assumption that it must be favourable for the animal practicing it under current conditions — are as a result very often ill-posed. With this framework in place, the reader is invited to consider some foci of current debates concerning the relevance of evolutionary considerations to topics of particular interest to contemporary moral philosophers, including altruism, reproductive strategies, and the control of aggression.

3 THE EVOLUTION OF ALTRUISM

The evolution of altruism presents a raft of theoretical and conceptual problems. Altruism and its relatives — co-operativeness, a disposition to fairness and promise-keeping even under circumstances unfavourable to the agent, beneficence, and benevolence — have traditionally occupied pride of place in moral theory. To what extent can these dispositions be explained as adaptations and to what extent are they maintained by other, non-genetic replicatory systems? The observation that many forms of human behaviour can reproduce themselves through the mediation of beliefs and attitudes and spread through a population and its descendants without contributing to fitness led Dawkins to introduce a separate, ideational system of replicators, which he called "memes" [Dawkins, 1976, 206ff.]. The position ascribed to Huxley is that human moral behaviour, especially the value it attaches to peace, justice, and fairness is entirely memetic; increasingly however, this position has been questioned.

It is easy to show that the theory of rational choice does not predict rewards for fair, altruistic, co-operative agents, but the introduction of "replicator dynamics" in which the payoff is not in dollars or well-being but in copies of oneself, changes the picture in some surprising ways. Sober and Wilson characterize altruism (in the sense relevant to evolutionary theory) not simply as helping behaviour, or even as helping behaviour involving certain risks to the altruist, but as "behaviour ... [that] involves a fitness cost to the donor and confers a fitness benefit on the recipient" and they propose an evolutionary explanation for it. On initial inspection, it seems logically impossible that such behaviour could be manifested in an organism let alone explained: how can a trait that renders each of its possessors likely to have fewer than the average number of offspring escape being driven to extinction? To solve their paradox, Sober and Wilson revert to the Darwinian hypothesis that altruism can be accounted for by competition amongst groups [Sober and Wilson, 1998]. Groups without any altruists die off; those with some proportion of altruists persist. The counter-intuitive aspects of the Sober–Wilson hypothesis, which does not satisfactorily explain how a low-reproducing altruist can invade a population of egoists in the first place, nor why an altruistic population is not subject to invasion by high-reproducing egoists, are somewhat mitigated by the observation that individual organisms are groups too — federations of genes and their products. Certain combinations are fortunate ones, even if some of their

individual constituents would fare poorly in other-things-being equal competition with rival genes.

It has nevertheless been doubted that group-living organisms actually behave in ways that satisfy the exigent Sober-Wilson definition, and it is not clear that behaviour involving a fitness cost as opposed to "vernacular" altruism is in need of explanation. Many theorists have pointed out that helping behaviour and a disposition to fairness and co-operation can benefit the donor as well as the recipient to a sufficient extent that the donor's type is not eliminated from the population and permitting the evolution of polymorphisms — a range of behavioural tendencies ranging from extreme generosity to free-riding and advantage-taking. Egoism is not an evolutionary stable strategy (ESS), defined as a strategy that does well when it meets itself [Maynard Smith, 1982; Skyrms, 1996]. Further, a gene for altruism would increase its representation in a small breeding community in which most members are close relatives [Rosenberg, 1991] who help each other. Altruism and fairness can be considered as forms of behaviour that make the social environment more conducive to living and breeding for an entire population, and emotion-based mechanisms that punish egoism with exaggerated retaliation might be expected to evolve in tandem with egoism, thereby limiting its effects [Trivers, 1971; Frank, 1989]. Finally, sporadic self-sacrificing behaviour that seems evolutionarily senseless, such as throwing oneself onto a hand grenade to save a comrade, or jumping into an icy river when one cannot swim to rescue a child, may represent an inadvertent overreaction of the human nervous system to an irresistible stimulus. The propensity to perform such actions in rare cases need not receive a direct evolutionary explanation. In some cases, especially where the reaction is instantaneous, it might be observed to resemble the immune system overreactions that sometimes kill the organism the system is adapted to preserving. In other cases, such as the devotion to a life of service by monks in a monastery or by a Mother Teresa or an Albert Schweitzer, the behaviour in question is clearly motivated by abstract ideals transmitted in revered texts and documents and unavailable to our nonliterate ancestors, who can be ascribed at best a latent disposition, present in at least some of them, to have self-sacrificing behaviour elicited by the appropriate ideational stimuli.

Given the range of behaviour possible for us, the provision of a satisfactory explanation for the evolution of some degree of altruism in group-living animals including humans does not settle the question how benevolent, co-operative, and fair we ought to be in the various situations that face us — how much to give to charity for example, whether to co-operate on some occasion or opt out of an existing co-operative arrangement, or whether to capitalize on a rival's weakness or ignorance. Yet the discussion of the evolution of altruism has had a remarkable effect on the construction of the image of human nature that furnishes the cultural background to contemporary moral debate. The error in taking the individual organism to be the privileged and unique unit of selection, the assumption that this individual was by nature selfish, and that the chief difference between man and animal in this respect was that man was motivated by considerations of

rational choice, is increasingly clear. Altruism co-exists with self-favouring and kin-favouring tendencies, automatic and emotional mechanisms co-exist with rational calculation, the latter depending heavily on social learning and formal instruction. The premise that human beings are basically selfish never actually licensed the inference that our economic systems ought to give free rein to selfishness, but it was assuredly taken to do so in many textbooks of economics. Depriving proponents of that view of their descriptive premise can accordingly stimulate or even force a re-evaluation of existing norms. Similar remarks apply to the discovery that strategies for pacification, conflict mediation and reconciliation are prominent in our near primate relatives [Flack and DeWaal, 2000]. The inference that major conflicts are best resolved by force of arms because that is the human way to do things is not only unlicensed, but ill-founded. The recognition that we may be endowed with an innate responsiveness to peace-making gestures, provided suitable signals are forthcoming from our opponent, can only encourage moralists who hope for a less violent world.

4 REPRODUCTIVE STRATEGIES AND THE TWO NORMS THEORY

The theory of evolution posits different dispositions, competencies, and preferences for males and females. Some theorists offer to derive the necessity of female subordination and exclusion from challenging, honourable, and lucrative pursuits from these differences. Interestingly, in much of the literature on Evolutionary Ethics, practices such as racial segregation and discrimination are regarded as the products of arbitrary prejudice, based in ignorance and harmful to persons and the community, while practices such as occupational segregation by sex are argued to be scientifically well-grounded and overall good for society, on the grounds that men and women are intrinsically more different from each other than members of different ethnic populations. The notion that human beings are a species that has evolved in such a way that its females lack the physical, cognitive and emotional resources that would enable them to attain male levels of political, intellectual and artistic expression, or that the species has evolved in such a way as to deny women opportunities for such expression, and that this state of affairs is desirable or inevitable, needs in any event critical examination.

Humans are physically dimorphic, though less so than are many ape species: gorilla males are twice as large as gorilla females, for example, and human males have lost the large canines possessed by male apes, suggesting that encephalisation is associated with a less combative mode of life. Human females are about 90% as tall as males on average, weigh about 80% as much, and have 30% less upper-body strength. Women have more fat and less muscle, differently structured immune systems, and accordingly different susceptibilities to disease. Hormones and neurotransmitters are secreted in different proportions. It is commonly assumed that not only reproductive efficiency but sexual selection has played some role in shaping the anatomies and appearances of the two sexes [Darwin, 1990; Cronin, 1991]. Cross-cultural differences between men and women that are generally agreed on in-

clude women's greater emotionality and empathy, women's greater predisposition to form intimate social relationships and lesser tolerance for physical pain, men's greater liability to anger, and more robust sex drive [Baumeister, 2000]. "Men are more likely to compete with one another for status using violence or occupational achievement", Stephen Pinker observes, "women more likely to use derogation and other forms of verbal aggression. ... Women are more attentive to their infants' everyday cries ... And men and women differ in their patterns of sexual jealousy, their mate preferences, and their incentives to philander" [Pinker, 2002, 345]. The two norms theory, as characterized by its critic Julia Annas [Annas, 1993], posits two different appropriate modes of life and customary occupations for, and two sets of standards by which to judge moral virtue, for men and women.

The existence of two different modes of life for the two sexes is noted by E.O. Wilson. "History records not a single culture in which women have controlled the political and economic lives of men ... Men have traditionally assumed the positions of chieftains, shamans, judges and warriors. Their modern technocratic counterparts rule the industrial states and head the corporations and churches" [Wilson, 1978, 128]. David Barash observes "There is no society, historically, or in recent times, in which women have not borne the primary responsibility for child care. ... In all societies, men do men things and women are left holding the babies ...." [Barash, 1979, 108]. Though the data set employed by these writers is mainly historical, the explanation is not: both Wilson and Barash suggest that men's political and cultural pre-eminence supervenes on basic emotional and attitudinal differences between the sexes, differences that reflect in turn their different reproductive strategies.

These differences are hypothesized to stem from the much larger investment women — like most mammals — make in direct child care, and in the shorter interval in which they can reproduce. Since male and female intelligence, creativity, task-persistence, and so on are not very different when measured objectively [Maccoby and Jacklin, 1975], the suggestion that men's and women's reproductive strategies offer a scientifically better and deeper explanation of women's lower degree of participation in influential, status-conferring activities than rival hypotheses citing arbitrary discrimination, malice, or cognitive or physical deficits has considerable plausibility. Pinker sums up the notion of differential parental investment and its relation to psychology as follows:

> Many of the psychological differences between the sexes are exactly what an evolutionary biologist who knew only their physical differences would predict. Throughout the animal kingdom, when the female has to invest more calories and risk in each offspring (in the case of mammals through pregnancy and nursing) she also invests more in nurturing the offspring after birth, since it is more costly for a female to replace a child than for a male to replace one. The difference in investment is accompanied by a greater competition among males over opportunities to mate, since mating with many partners is more likely to multiply the offspring of a male than the number of offspring of a

female [Pinker, 2002, 346].

Barash argues that "Males tend to achieve fitness by making themselves as attractive as possible to females, then rely largely on the females to take it from there. Often they compete with other males, either for direct access to mates or for access to resources which help them acquire mates." Men are hypothesized to have more time available for mating effort and to rely on mating effort, including competition and display, rather than on paternal care as their major form of reproductive investment. To be sure, one can always cite exceptions to generalizations such as "women are more emotional than men are". However, the existence of exceptions is not a powerful objection against the claim that statistically-determinable differences between the sexes, when large, can both explain and justify the instantiation of two basic norms. stronger objections are needed and are availab le.

Though Pinker acknowledges some role for discrimination, he is persuaded that women's socio-political deficits flow to a significant degree from their own preferences, [ibid., 352–3]. Though he does not suggest, as E. O. Wilson did, that tampering with the balance of power by fiat might be inadvisable, he appears to agree with Wilson that equality of socio-economic outcomes for men and women is not a reasonable social goal, as long as "People vary in traits relevant to employment ... Given all the evidence for sex differences... the statistical distributions for men and women in ... strengths and tastes are unlikely to be identical. If one now matches the distribution of traits for men and for women with the distribution of the demands of the jobs in the economy, the chance that the proportion of men and women in each profession will be identical, or that the mean salary of men and women will be identical, is very close to zero". Pinker quotes with approval the view of L. S. Gottfredson: "If you keep using gender parity as your measure of social justice, it means you will have to keep many men and women out of the work they like best and push them into work they don't like" [ibid., 359].

The evolutionary argument for two norms can be stated as follows:

> Evolutionary Argument for Two Norms: Because of their different specialisations in reproduction, men and women find different categories of experience pleasurable and aversive. The realisation of a single norm with regard to typical patterns of behaviour (employment, childrearing, sexual behaviour) would be experienced as oppressive for at least one sex and possibly for both. The gains to be achieved in terms of overall human preference satisfaction from a uniform set of expectations are nonexistent or small. Hence the realization of two norms is not only inevitable or overwhelmingly likely, but morally permitted or indeed obligatory.

The fully developed two norms system thus incorporates a set of ideals, permissions, and obligations. Its chief features are that women are guarded with the help of a system of workplace segregation and confinement in the home. Being of a retiring, dependent, and child-oriented nature, they are presumed not to experience this as deprivation, and political activity, the exercise of executive authority,

and cultural accomplishment are left to men who are held to express their competencies and intrinsic competitiveness by these means. Statistical equality — the equal participation of women in politics and governance, the full spectrum of economically productive activities, and academics and the arts — is considered to be a false ideal since women allegedly do not have time, the motivation, or the appropriate character traits, as imprinted in them by natural selection in competition with other women, for full participation. This position needs critical examination.

5 REPRODUCTIVE STRATEGIES AND SEX ROLES

"Males", Matt Ridley remarks, "usually compete for access to females, rather than vice-versa. There are good evolutionary reasons for this, and there are clear evolutionary consequences, too; for instance men are more aggressive than women" [Ridley, 1994, 13]. In a chapter of *The Red Queen* entitled "Polygamy and the Nature of Man", he aims to illuminate the subject of human sexual behaviour by references to peacocks and peahens as follows: "She will do all the work ... Every time [a peacock] seduces a fresh female, he wins the jackpot of her investment in his sons and daughters .... In more human terms, men can father another child just about any time they copulate with a different woman, whereas women can bear the child of only one man at a time" [Ridley, 1994, 172–3]. Though the chapter ends with the observation that, in our species, women do in fact compete with men, suggesting that some features or qualities of men are a scarce resource where women are concerned — the initial passage suggests, not only that women should be grateful that they are not as badly off as peahens, but that that men are foregoing a "jackpot", by not trying to seduce as many women as they can.

On one popular view, men are naturally polygamous, insofar as their fitness depends chiefly on the number of women they are able to impregnate, while women are naturally monogamous, insofar as their fitness depends chiefly on the number of offspring they can raise to maturity. Lifelong, sexually exclusive marriage is hypothesized to be a female preference, and either polygamy or promiscuous bachelor independence a male preference. To achieve their goals of comfortable and effective child-rearing, women allegedly accept a service role vis-a-vis a particular man and tolerate his infidelity. To achieve their own goals of seduction and reproduction, men in turn allegedly assume a provider role towards a particular woman but take advantage of additional "mating opportunities". Human males sacrifice some degree of sexual freedom and assume some degree of responsibility, in exchange for which females assume a somewhat wearying caretaking and maintenance role. Each sex is hypothesized to get most of what it wants under this system, though stresses are acknowledged as inevitable: females are alleged to fear desertion, males, cuckoldry.

The image of human nature depicted here is, however, misleading. It is true that while both men and women have the same high standards for long-term partners, valuing equally such traits as good health, kindness, intelligence, and creativity, women's acceptability threshold for short term sexual partners is considerably

higher than men's [Kenrick *et al.*, 1990]; further, there are good evolutionary reasons for this: namely that, as Trivers remarks, "errors in mate selection are generally more expensive to females than to males" [Trivers, 1972]. However, the role of aggression in the formation of human relationships is not distinguished in Ridley's comment from its often futile or tragic role in attempts at relationship maintenance, nor is the vernacular sense of the term competition effectively distinguished from the technical sense, rendering his remarks somewhat misleading. Aggression plays at best a ritualistic role in human mate-selection (tournaments, etc.), and while men initiate, or decline to initiate socio-legal marriage proposals in most societies, the preferences of women are decisive at very early stages of courtship, rendering aggression somewhat beside the point. Women appear to have an innate or easily learned mastery of "courtship de-escalation strategies" [Moore, 1998], and female dominance in this realm has been observed in elephant seal, mice, fish, rats, gorillas, monkeys and birds [ibid.].

Women's capacity for attachment meanwhile is a salient feature of the phenotype, evidently distinguishing women from peahens. Its adaptive significance is unclear. Love has been hypothesized as a form of stop! command that "may facilitate rapid decision by putting strong limits on the search for information or alternatives," insofar as checking out many potential partners wastes time and the best discarded candidates are unlikely to remain available [Gigerenzer *et al.*, 1999, 292, 363]. But it may also be the case that, as with birdsong, nature has invented the keys needed to unlock certain locks. Dawkins's suggestion that the effects animals exert over each other serve their own interests, not that of their helplessly mesmerized targets, is worth considering in this connection.

"In principle", as many evolutionary theorists never tire of pointing out, a man can father thousands of children — and neglect them too. A man can have — and some men with extraordinary coercive powers, economic leverage, or a flair for deception do have — two or more full-term children whose birthdays are nine months apart or less, whereas a woman cannot and no women do. An average man is, however, statistically no more likely to father a child every time he takes a different woman to bed than every time he takes the same woman. Since he is in competition with other men, many of his attempted seductions will not succeed, and most of his successful seductions will not result in a child. As astonishing as it might seem, the average number of offspring an average man can expect to father is exactly the same as the average number of offspring an average woman can expect to bear! Under conditions probably obtaining in the Early Adaptive Environment — a low fat diet, consumption of fertility-depressing phytoestrogens, and prolonged lactation — this number was probably 4-5. The variance between men in this respect may have been considerable, but we do not know how great it was in earlier environments. Nor do we know how much of this variance depended on *behavioural* strategies, including promiscuous or nonpromiscuous sexual behaviour, paternal and spousal care, the provision of food, or status-seeking, as opposed to *nonbehavioural* but potentially equally effective biological strategies such as possessing good looks, or a good immune system, or other such characteristics. In

the absence of this information, it is hard to tell what behavioural strategies were selected for. The assumption that women derive a benefit from their social association with men is clearly inconsistent meanwhile with the assumption that men rely on women to "take it from there" once they are impregnated. The same point holds for women; we do not know how much of the variance in female reproductive success depended on individual behavioural variables, including promiscuity, fidelity, maternal care, caretaking of males, or aggression towards other females or their offspring. Nor do we know how much depended on behavioural vs. nonbehavioural strategies including the possession of good looks, intelligence, or a good immune system.

Studies of revealed preference shed less light on the question of male and female traits selected for than is generally supposed. In a large cross-cultural study of more than 10,000 individuals, David Buss found that, while both sexes overwhelmingly regarded intelligence and kindness as their main criteria in selecting mates, men preferred young and physically attractive mates, whereas women attached less importance to age and physical appearance and more to economic advantage [Buss, 1989]. This finding might seem to license the conclusion that women who allocate time and effort to establishing lucrative careers will place themselves at a disadvantage relative to women who marry young and allocate time and effort to beautification, and that the former ought to be dissuaded from career pathways that are advantageous to men. A similar conclusion might be drawn from the much-discussed male preference for particular female waist-to-hip (WHR) ratios, which men are able to discriminate very accurately, and which are hypothesized to be reliable signals of fecundity [Singh, 2002]. It might seem rational, in light of this datum, for men and women to allocate their efforts differently if their main goal in life is to attract a large pool of potential mates or to have as many children as possible.

It is unlikely, however, that women evolved chiefly to give reliable signs of fecundity and to devote the majority of their efforts to giving unreliable signs of fecundity by way of time-consuming beautification to the exclusion of other forms of achievement. Buss's research has been criticized as unilluminating, since the respondents were all university educated and the cultures studied were those in which a woman's status and well-being depended significantly on her husband's income [Boyd and Silk, 1997, 645–7]. In the early adaptative environment, female status, whatever it depended upon, could not have depended upon this. Moreover, the rigours of the early adaptative environment, the need for social intelligence, vigilance, competence at provisioning, and teaching ability on the part of successful mothers, make it unlikely, as Hrdy points out, that the Pleistocene woman who relied on her beauty to pull her offspring through left many descendants [Hrdy, 1999, 24]. It also renders it unlikely that men who selected mates chiefly on the basis of the WHR did so either.

Women's alleged preference for older and more established mates, it should be noted, does not translate either into a direct preference for such persons as sexual partners, as opposed to protectors and nurturers, nor into a direct preference

for high-earners. While meat in large quantities can only be supplied by male hunters, women in hunting and gathering societies meet most of the nutritional needs of their communities, by gathering, digging, trapping, and fishing. Only in large, hierarchically-organized societies in which women's economic activities are formally constrained does earned, appropriated, and inherited wealth become a scarce and necessary good, to obtain a share of which women will make evident sacrifices. In any case, it would be difficult to mount an argument that social institutions and educational systems ought to assist women in pursing the goal of an excellent WHR, rather than in their intellectual development or maximal use of talents.

6 MORALITY AND THE MARITAL COMPROMISE

The notion that human marriage is a compromise between the male preference for polygamy or free-ranging bachelorhood (as opposed as these strategies are) and the more uniform female preference for lifelong monogamous commitment is not well-supported by observation. Rather, in addition to patterns of sexual receptivity that correspond to lunar months and to the seasons, there appear to be longer cycles of emotional receptivity determining the timing of human association and defection. A spontaneously arising and collapsing pair bond of four to seven years duration, punctuated by opportunistic excursions, appears to be natural for both sexes, with women's selectivity in the latter regard somewhat higher than men's, as the evolutionary theorist would predict. The wastefulness of relentless and futile seduction efforts by males, together with the sacrifices for both sexes of strict monogamy, suggest that a mixed strategy will be realized, either, as Trivers suggested, as a polymorphism within the group, with various temperaments representing a variety of strategies, none of which goes to fixation, or by shifts of habit over the life-cycle of the individual, or in response to surrounding conditions. As Kitcher remarks, "Males and females can be expected to play highly complicated conditional strategies whenever evolution has equipped them with the cognitive capacities required by these strategies" [Kitcher, 1985, 173]. Contrary to the claim that females overwhelmingly prefer monogamy, it is obvious that females are "fickle" in the sense that they are likely to revise their initial estimates of mate quality on the basis of longer experience. Many novels and stories deal with the profound effects on individuals and on the social fabric of such re-evaluations. Further, as men are known to defect, die, and kill each other, a rigid monogamy strategy would be easily invaded by a more flexible temperament. While men would not tolerate harem-like conditions, DNA analysis has demonstrated that covert polyandry is more extensive than previously suspected.

The notion that women object less to physical infidelity on the part of pair-bonded mates than to "emotional" infidelity and the threat of abandonment because they do not risk raising children who are not their own has been questioned for both its factual and its explanatory component [Harris, 2004]. Jealousy seems not to be very differently structured in the two sexes, assuredly not differently

enough to support a double standard. However, sexual jealousy is a leading cause of homicide in present-day hunter-gatherer societies [Boehm 2000, 93-4]. The imposition of harsh punitive codes in larger societies with an effective legal structure is doubtless an attempt to reduce social conflict and forestall lethal outcomes. Such codes target women's behaviour presumably under the assumption that women are easier to control than men are. From a moral point of view, however, despite the need for effective measures to prevent homicide, such measures as stoning and ostracism are unacceptable, as is the extra moral burden born by females in this department.

Marriage — a form of bond that may, but need not be supported by emotional attachment and exclusive sexual preference — by contrast with the natural pair-bond, appears to be an institutional solution to a number of problems that did not manifest themselves in the early adaptative environment. These include the social usefulness of the exchange of women in cementing relations with neighbouring tribes, as increasing populations came to encroach upon their neighbours; the inability of women, in later conditions of urbanization, at all stages of their lives, reliably to provide for their nourishment and that of their offspring without male assistance; the inability of men to compete effectively in the public realm in the absence of a maintenance staff to provide meals, cleaning, and clothing, and the need for clear lines of inheritance in societies in which wealth is accumulated and bequeathed by men. These "unnatural" aspects of marriage raise important questions for the moralist. Would women's full participation in the labour force, or different systems of inheritance, render marriage obsolete? Are there reasons connected with human well-being for encouraging and providing incentives for marriage?

The tendency of the natural pair bond to degrade can be countered, wherever sentiment is insufficient, by constraining forces or appropriate ideation. Whether it is worthwhile on a social or individual level to invest effort in marriage depends on the value attached to its various functions and aims. These include the raising of children known to be one's own, the efficient division of labour, the pooling of resources, the sharing of a history, and safe sex in the psychological as well as the microbiological senses. Modern societies are notoriously conflicted in that they are inclined to support the institution of marriage by e.g., imposing penalties on defectors and making their exit as complicated and unpleasant as possible, at the same time as they attempt to show compassion for unhappily trapped individuals by easing on the price of exit. The proper balance is a matter for debate, but the suggestion by some authors that whimsicality and intolerance [Cere, 2001] must lie at the heart of all marital breakups does not take into account either the capacity of humans to learn about other people from experience, or their capacity for making others near them miserable.

7 ENHANCEMENT NORMS AND CONTROL NORMS

An anthropologist studying human social systems might observe that human beings employ both what might be called "control norms" and what might be called "enhancement norms", and that many social rules are fusions of the two types. "Control norms" target natural dispositions as needing to have their expression blocked. Examples of control norms are the proscriptions on the murder of annoying persons, or useless and possibly dangerous strangers, lying for personal gain, and the stealing of valued resources from persons helpless to defend them. Enhancement norms, by contrast, have a quasi-aesthetic character. They elevate observed tendencies to the status of obligations, or demand that actions be practiced with special intensity, or that attributes be exaggerated for effect. Men's wristwatches, for example, are standardly twice the size of women's, though men do not have wrists twice as thick, but only somewhat thicker.

Both the ideal of monogamous marriage and the special exemptions offered by the double standard qualify as enhancement norms, for they transform tendencies — the pair-bonding instinct, women's greater and men's lesser selectivity — into obligations and permissions. They also function as control norms, proscribing behaviour that would occur more frequently in the absence of the norm and its mechanisms of enforcement. The norm that mandates that mothers abandon economic activity outside the household to care for their own children is an enhancement norm, since all primate females are somewhat encumbered by their young and by maternal care. But it is also a control norm in human societies, for the primate mother does not experience such severe restrictions on her mobility, and may receive considerable help from her relatives and from young females desiring practice with infants, while she engages in foraging or socializing [Hrdy, 1999, 497ff.]

To evaluate the claim that the division of labour, social roles, and entitlements by sex is morally legitimate, we would have to know how to determine, measure, and compare the morally relevant gains and losses of various possible systems. The suppression or alteration of psychological tendencies, such as women's diffidence and modesty, or men's general inability to feel seriously intimidated by women, the additional educational resources needed to develop women's interests in the most "inhuman" branches of science and technology, the large social investment that would be required to provide child care equivalent to that historically provided by mothers, all represent large social costs. The gains, however, measured in terms of the overall appearance and functioning of the present world and the realization of ideals such as development of autonomy in individuals and the elimination of all residual forms of slavery are also considerable. The vernacular possibilities of parasitism and obsolescence, understood in a sense analogous to Dawkins's, make it implausible to suppose that women's more restricted lives are in the interests of the persons living them. Like the addictive behaviour of the cuckoo's foster-parent, some typical-seeming aspects of women's behaviour may be triggered by irresistible stimuli without stemming from their more authentic drives and inter-

ests. Or, under modern conditions, women's evolutionarily-developed preferences may no longer correspond to their all things considered interests, their actual competencies, or the benefit their greater participation would in fact bring in terms of the peacefulness, health, political stability, and aesthetics of the common world.

While Evolutionary Ethicists tend to regard convergence towards a single norm as an example of the attempted override of a natural system that is bound to provoke confusion and resentment, defenders of the single norm insist that the high degree of polarization of modern societies is itself an example of override by non-biological replicating memes. For it is a feature of the human phenotype that it is capable of generating and maintaining institutions, elements of the extended phenotype, that serve some interests at the expense of others. A better understanding of human nature promises, in Pinker's very appropriate terms, "a naturalness in human relationships, encouraging us to treat people in terms of how they do feel rather than how some theory says they ought to feel". Certainly, additional increments of effort would be required to e.g., teach or acculturate women to smile only as much as men do, and additional increments of effort (or perhaps just large doses of drugs) would be required to e.g., teach or acculturate men to care for infants with the same success rate in keeping them alive as women. But the notion that the costs of greater equality require these sorts of adjustments, and that the costs of more relevant adjustments are unpayable, is untenable.

It is useful to remember that psychological dimorphism is constrained by the same features as physical dimorphism and that, embryologically, males are modifications of the underlying female type. The remarkable shift of attitudes within a single generation in Europe and North America with respect to direct paternal care is significant in this connection, in showing that Barash's casual view that males "rely largely on the females to take it from there" ignores certain latent potentials that the naturalist Réamur noticed in the 18th century and that Darwin commented upon in 1875. "A cock, by being long confined in solitude and darkness, can be taught to take charge of young chickens; he then utters a peculiar cry, and retains during his whole life this newly acquired maternal instinct." [Darwin, 1875, II: 26–7] That such drastic priming measures are not needed in human beings [Hrdy, 1999, 205ff.] is indicated by the fact that this shift, though triggered by abstract philosophical considerations of equality, seems in many cases to reflect the revealed preferences of men, preferences that could not be expressed without shame in earlier two norms systems.

To summarize, the study of reproductive strategies has the potential to shed a great deal of light on moral and social issues, ranging from childcare and the division of labour, to divorce and inheritance laws, and the function of education in both enhancing existing competencies and in compensating for or controlling deficiencies and imprudent tendencies. Yet it is the sector of evolutionary theory most liable to confusion and distortion. It resembles in this respect earlier discourses on sexuality [Foucault, 1988, 54 et passim; Kitcher, 1985]. It is striking, for example, how little effort is often made to think through the consequences of the limited and obscurely signaled fertility of human females by contrast with that

of most mammals, and how inconsistent with the premise that feminine beauty is the chief qualification in the human mating system the premise that most of the work of childcare is performed by women is. The notion that conflict pervades the relations between the sexes because polygyny is natural for men and monogamy for women oversimplifies a complicated dynamic, and while fantasy is a source of revealed preferences, the frequent intrusion of harem imagery into discussions of human propensities is at odds with the observation that harem-formation and maintenance is not practical in hunter-gatherer societies and is unlikely to represent the original mating system of human beings.

It seemed evident to the ancients that nature had intended women for childbearing and maintenance activities, and that women were to be assessed in terms of those functions, and that nature had intended men for the participation in the legal, intellectual, military, and political institutions that constituted the rest of human life. Contemporary moral theory, by contrast, operates with notions of agency that give no weight at all to constraints imposed by sex, with the somewhat paradoxical result that observed discrepancies in the overall well-being of men vs. the overall well-being of women emerge as morally irrelevant. The two norms theory thus seems established by fiat in traditional moral theory and by default in modern. Evolutionary Ethics is capable in principle of offering a third and distinct perspective, an alternative to universalist indifference, as well as to the absurdities of cosmological deductions of men's and women's roles. The plain fact is that nature "intends" both sexes for reproduction.

8 THE OPEN QUESTION ARGUMENT

The foregoing discussion suggests that evolutionary theory has the power to illuminate many of the conflicts and struggles experienced by individuals as well as the difficulties involved in framing appropriate social institutions. But can it do more than illuminate problems? Can evolutionary theory also point the way towards solutions?

The traditionalist acknowledges that humans evolved from nonhuman primate ancestors, that they possess psychological and behavioural traits that, like their morphological and physiological traits, have been shaped by environmental pressures and competition with conspecifics, and that some of our current traits are shared with successful ancestors, while others represent species-specific innovations. He may readily agree that the behaviour interesting to the evolutionary theorist — aggression and pacification, altruistic and co-operative behaviour, parental care, parent-offspring conflict, the division of labour, acquisition and sharing, and the struggle for status and dominance — is the same as the behaviour interesting to the moralist. He may even agree that many features of human social life that we are antecedently disposed to regard as good, such as co-operation and parental devotion, seem to lend themselves to explanation by reference to competition and selective pressures. The theory of evolution can probably explain why we are suspicious of people with different facial features (they might be hostile and

dangerous); like sweet and salty foods (they contain rare and needed nutrients); and even why psychopathic leaders can dominate entire countries (we are disposed to trust people who appear confident). Natural selection can also presumably explain why we like stimulants and depressants such as alcohol, caffeine and tobacco; this liking is a byproduct of chemical receptors evolved for other vital purposes. Parent-offspring conflict in adolescence, and the intergenerational hostility that results not only in the complex sanctions placed on young people's desires, but in young men being sent to war, to the distress of their mothers, by older men, has a rather obvious potential explanation in terms of reproductive competition. Even the liking for risky behaviour on the part of young men, the cause of so many social problems, may have an explanation in these terms. (Such behaviour is often said to stem from a desire to impress women, who are presumably impressed, but, recalling Dawkins's observation that stereotypical behaviour is not always in the interests of the organism performing it, a more subtle explanation might cite peer pressure as a form of victimization intended to eliminate potential competitors from the breeding community.)

Nevertheless, the traditionalist will insist, the supposition that some human traits were adaptive in the past, or gave their possessors a reproductive edge, and that some patterns of behaviour represent evolutionarily stable strategies, or correspond to various types of durable polymorphism, can have no implications for prescriptive moral theory. According to the terms of G. E. Moore's Open Question Argument [Moore, 1903, Ch. 1], we are always in a position to say "I appreciate that behaviour pattern X represents an ESS, but why is X morally good?" or "I appreciate that behaviour pattern Y is fitness-inducing and predicts a larger future lineage than its competitor patterns, but does this mean that Y is morally permissible?" Why does it matter, the skeptic will ask, whether some form of behaviour that is subject to moral regulation, whether aggression, lust, lying, selfishness or neglect, is an element of a human survival-and-reproduction-strategy or the result of interaction between a talking serpent and Adam and Eve? When it is argued that because "[t]he Darwinian does not expect you to feel guilt about spending money on your children rather than sending it all to Oxfam", that you are right not to feel guilt and that you are not obliged to send it all to Oxfam [Ruse, 1986, 239], it is evident that another premise is needed, though the conclusion is clearly true.

To bridge the gap between "is" and "ought" and establish the very possibility of Evolutionary Ethics, some guiding principle is required, and the best candidate would seem to be the following:

> Is-to-Ought-Principle: Since behaviour patterns in a species that have arisen as a result of selective pressures are apt to be pleasurable and satisfying to the organisms engaging in them, supported by redundant mechanisms, and overprogrammed or difficult to extinguish, one ought not try to suppress behaviour that can be shown to have arisen in the course of evolution, unless there are strong reasons for doing so. Either the attempt will likely produce stress if repression is successful, or it will

> consume resources to no avail if it is not. Further, if some alternative practice or form of social organization would permit people to act in ways that were more natural, hence more pleasurable and satisfying, then it is obligatory to introduce them unless there are strong reasons for not altering the practice.

This view of the role of evolutionary theory in practical ethics falls well short of the enthusiastic suggestion of evolutionary theorists, who insist that, since there are biological, genetic and natural components to our behaviour, we should organise the social world to match tendencies.

The sense of "matching" envisioned is in any case unclear. Virtually all moral theorists regard some forms of behaviour observed in our primate ancestors and probably inherited from them, including reciprocity, empathy, loyalty, and reconciliation, as deserving praise and reinforcement, at least in some contexts. But the reason why our institutions should "match" our tendencies in this regard is not that they have evolved, but because we think on independent grounds that they are good. Reciprocally, the notion that we should "set up" a social world that gives free rein to aggression, racism, sexism, deceitfulness, and all-around parochialism because they have evolved has nothing to recommend it. The task for the moral theorist is to distinguish between, on one hand, natural behaviour whose suppression is or would be cruel, arbitrary, and pointless, and unnatural behaviour that it is futile or excessively costly to try to inculcate, and, on the other hand, natural behaviour that deserves suppression and unnatural behaviour that needs to be inculcated, even if it is somewhat costly to do so. To base normative ethical theory on human wants and needs — whether these are the revealed preferences of social science or are hypothesized as old adaptations — is basically sound, but the devil is, as usual, in the details. We live simultaneously within a number of different mental systems whose contents do not always mesh, and cherished plans and goals are often in conflict with one another. There are inconsistent preferences, perverse preferences, short-sighted, imprudent preferences, immoral preferences, and conflicting and incommensurable preferences. "The maximal satisfaction of the preference-set of human beings" is not a well-defined concept, or something it is sensible to aim at. Human life is, in Midgley's terms "a rough but tolerable equilibrium" [Midgely, 1978, 282], in which we mostly try for local and speedy improvement.

But if Darwin was correct, one might wonder, how can there be objective moral truths or objectively morally better states of the world? Are we not driven to a metaethical form of what Alex Rosenberg and Tamler Sommers term "Darwinian nihilism", [Rosenberg and Sommers, 2003] as articulated by E.O. Wilson and Michael Ruse: "Ethics in an important sense, ethics as we understand it, is an illusion fobbed off on us by our genes to get us to co-operate. Furthermore, the way our biology enforces its ends is by making us think that there is an objective higher code to which we are all subject" [Wilson and Ruse, 1989, 51]. Fallacious ideation, it is tempting to speculate, such as a belief in punishing and preserving gods, and an age-old tendency to confuse kings with gods, is probably a condition

of human societies being governable, especially large societies. Darwinian nihilism is nevertheless not a coherent philosophical position. In fact, social animals will co-operate whether they have a system of morality or not, according to game-theoretic analysis, because co-operation enhances fitness, or, more precisely, non-co-operation can put an animal at a disadvantage. In thinking there is a higher code, we have the thought that some nonprudential behaviour — including often the refusal to co-operate, with, for example, an unjust regime, or in a war — is more noble, valuable, obligatory, etc. than prudential behaviour.

Since anyone's pattern of behaviour lends itself to description and evaluation in moral terms, and since, in conscious creatures, behaviour patterns of social complexity generally require reference to beliefs, the formation of moral beliefs is a human universal, and it is difficult for humans not to be concerned with the contents of one another's beliefs as well as their overt behaviour. One might as well argue that the contents of cookbooks are illusions fobbed off on us to get us to eat as argue that the contents of ethics books are illusions fobbed off on us to get us to co-operate. We have evolved, not only the disposition for unselfish behaviour, which makes morality possible, but, equally importantly, the cognitive abilities needed to represent merely hypothetical states of affairs, to attach evaluative predicates to these representations, and to prefer well-reasoned argument to poorly reasoned argument, making moral theorizing and its application possible. Morality cannot therefore be "imaginary". Just as more critical and sophisticated ideas about the basis of political obligation have succeeded divine right theories without bringing about a collapse of political authority, more critical and sophisticated views concerning the meaning and truth-status of moral judgments have succeeded ancient divine command theories, without ethics losing its regulatory importance.

As noted earlier, it is often claimed that Evolutionary Ethics contains an intrinsic bias towards social inequality, and it can be difficult to separate the implications insinuated by its practitioners from the implications of the data they cite. Most evolutionary ethicists recognize in principle that some human dispositions, though "natural," ought, for moral reason, to be controlled: Thus Barash in *The Whispering Within*:

> If sociobiology is correct, we've got to be carefully taught not to hate others who are different from ourselves, because it may be our biological predisposition do so. If evolution does incline us to a degree of racial bigotry, that certainly does not mean that such inclinations are justified ... Perhaps one lesson to be gained from sociobiology is that we must demand that our cultural institutions such as education and child-rearing make sure that we are "carefully taught" to love one another [Barash, 1979, 155].

The vigour with which they denounce racism or the persecution of homosexuals and insist on correction by culture makes the evident tolerance of sociobiologists for sexism very difficult to understand, and their tendency to trivialize revisionary

social movements as "political correctness" seems strangely misguided. Pinker argues that the denial of what he calls "human nature" has promoted cruel and unnatural childrearing regimes, parental guilt over children who turn out badly, urban planning that violates the human desire for natural light, ornament, and surroundings to scale, and the release of psychopaths under the illusion that they can be reformed by counseling [Pinker, 2002, xi]. It has also, one might add, promoted toleration of psychopaths under the illusion that they are great men, racial oppression under the illusion that marked differences in facial features and skin colour must signal marked differences in cognitive ability, the persecution of homosexuals, and cruel marital regimes.

To clarify the relationship between social inequalities that arise out of real, and therefore natural differences in human beings within certain social frameworks and the demands posited by philosophical morality, the Evolutionary Ethicist might be presented with the following Basic Factual Claim:

> Basic Factual Claim: There is a correlation, positive in some cases, negative in others, between being a member of a particular racial or national group, or of one sex rather than the other, and experiencing a better or worse life. Sex, race, and nationality have a predictive value with respect to well-being, not only with respect to income, prestige, executive authority, and so on.

An Evolutionary Ethicist would be ill-advised to reject the Basic Factual Claim, since there is abundant evidence that sex, ethnicity, and parentage influence one's likelihood of experiencing poverty, abandonment, domestic violence, isolation, poor health over the long term, and so on. Moreover, accepting it does not commit anyone to the normative claim that so-called statistical equality is a desideratum. The Evolutionary Ethicist must now decide whether to accept the following Basic Normative Principle.

> Basic Normative Principle: There should be no correlation, positive or negative, between being a member of a particular racial or national group, or of one sex rather than the other, and enjoying or suffering a better or worse outcome. Sex, race, and nationality *ought to have* no predictive value with respect to well-being — liberty, security, and enjoyment of life — whether or not they have a predictive value with respect to income, prestige, executive authority, and so on.

Note that an Evolutionary Ethicist who accepts this principle is still not committed to the claim that statistical equality is morally desirable or obligatory. For there may be no correlation between objective well-being and the possession of income, prestige, authority, and so on.

If the traditional moralist would be well served by acknowledging the validity of the Is-to-Ought-Principle, the evolutionary ethicist would be equally well served by acknowledging the validity of the Basic Normative Principle. It is difficult

knowing all that we moderns know to reject the latter Principle on the grounds that it represents an ignoble and unworthy ideal, though many notable moral philosophers of the past would most certainly have rejected it on precisely those grounds. But might one insist that the Basic Normative Principle is untenable in light of the Is-to-Ought-Principle, because it is just too hard, given the sorts of creatures we are, to prevent well-being from tracking sex, race, etc.? Moral rigorists have in turn rejected the Is-to-Ought-Principle, but their reasons for doing so outright seem exceedingly specious. One might instead adopt the more cautious position that we do not know whether the Basic Normative Principle ought to be upheld at the costs it would impose because we do not know how hard it would be for everyone — not just members of better off groups — to bring about or live in world in which well-being was uncorrelated with sex and race. We should therefore accept this principle for the moment as representing a valid ethical ideal — one that we do not know we can reach, but one that we have no particular reason to think we cannot reach.

In conclusion, the theory of evolution furnishes a kind of heuristic for the understanding and evaluation of institutions and practices, though it cannot evaluate them as directly as some of its proponents assume. The major contribution of biologists and evolutionary psychologists to the field of ethics consists in demystifying many morally problematic features of human life, such as the striving for status and conquest, belligerence, sexual subordination, and intra-familial conflict. These phenomena were in ancient times conceived in politico–religious terms as involving values, in epistemological and moral terms as involving conflicts between people who were right or superior, and those who were wrong or inferior. They were the interventions of rivalrous and emotional deities beyond human control who caused not only storms and earthquakes but war and civil and family disturbances. General systems of the world, like those of Plato and Aristotle, developed the notion that, throughout the cosmos, higher entities rule lower entities, that active is superior to passive, and that the goodness of a thing is related to its fulfillment of its intended function.

The new naturalism is accordingly distinct from Aristotelian naturalism. With demystification, opportunities arise for human rationality and sympathy and a sense of fair play to be applied to a wider range of phenomena than were historically seen as eligible. The notion of empire, much discussed in the earliest Darwinian literature, furnishes an example. Empires are phenomena that recur repeatedly in human history, and, while some features of their replication are obviously memetic — since imperialists assiduously study and imitate the actions of their imperialist predecessors — empires would presumably arise spontaneously even if imperialists did not know of one another's existence. In this respect, they can be considered as features of the extended human phenotype, as habitats of a partially symbolic but partially concrete sort built by organisms of a particular species that they like to inhabit. The propensities of humans appear however to depend in turn on political conditions. Remarkably, while humans exhibit a strong bias towards egalitarianism, conflict resolution and sharing in small groups where individuals

know each other well [Boehm 2000], in larger societies they gravitate towards autocratic or at least strongly hierarchical systems, militarism directed against outsiders, and marked social inequality. This flexibility is at once impressive and disturbing; it raises important questions for social policy and for the choice of social leaders. In any case, understanding the roots of political organization and the alternative forms than human life once assumed enables us to be appropriately skeptical in face of the noble language in which expansionist and inegalitarian policies are typically dressed. Democracy, world federalism, and disarmament movements are counteracting control norms that, at the same time, qualify as enhancements of the human propensity for reflective thought and pacific behaviour. Moral progress consists in adjustments of enhancement and control norms in ways that acknowledge the respect in which features of the contemporary environment render them obsolete or parasitical, or that show them up as based in ignorance, or as flowing from goals that reasonable human beings do not share with their genes.

The study of adaptiveness does not, on this view, directly posit norms or generate prescriptions. Rather, it contributes to the construction of a scientific image of human nature, as incorporating what Christopher Boehm describes as a set of "contradictory dispositions that generate practical ambivalences at the level of phenotype, ambivalences that help to structure life's practical decision dilemmas." [Boehm, 212] Evolutionary theory can alert us to what arrangements and practices individuals are likely to find oppressive and what behaviour it is likely to be difficult to change without considerable effort and imagination. This scientific image is more objective than the array of basically distorted images of human nature presented by fiction, historical literature, the daily newspaper, "common sense", and other informal sources of information about what people are like; to base social policy on what people *believe* about members of their species seems absurd, if one has the genuine option of basing it on what can be known, or at least surmised with confidence.

None of this implies that nothing can be demanded of humans that is not based in the proto-morality of our primate ancestors, or that the scientific image can generate any imperatives. The expansion of the neo-cortex in humans led to the emergence of cognitive and emotional capacities that expand both the capacity for advantage-taking, persecution, and organized cruelty, and, at the same time, for the extension of beneficence and concern well beyond a narrow circle of relatives and intimates. The ability to posit abstract ideals, to recall a distant past, and project a distant future in the imagination, to be impressed by the moral lessons of myths and stories and to try to emulate heroic conduct, to understand and empathize with or conversely to dissociate oneself from another person's goals, are capacities that animals lack. These capacities are species-specific and therefore heritable, even if the form in which they are developed and expressed varies from culture to culture. Nor do Evolutionary Ethicists have any basis for denying the Basic Normative Principle. If traditional ethicists, meanwhile, could be discouraged from rejecting the Is-to-Ought-Principle out of hand, conflict between the two

parties could be set aside in favour of more productive debate about particular social norms.

E.O. Wilson has written affectingly of what he terms biophilia, "the innate tendency to focus on life and lifelike processes" [Wilson, 1986]. Our species evolved amongst plants and animals, not in concrete apartment blocks and towers with windowless offices, illuminated by light of an unusual spectral composition, and there can be nothing objectionable in the suggestion that a restoration of original environmental conditions insofar as this is possible might promote a sense of well-being. Clearly, humans have chosen to live in ways that do not fully correspond to their evolutionarily-determined preferences, either because they have other, overriding preferences, or because their evolutionarily-determined ones are inconsistent with one another. As ergonomic design results in products more comfortable and convenient for human grips, gaits, and postures at the expense of decorative values, normative moral theory seeks out designs for living that are more comfortable and convenient for the specialized creatures we are, though we know, or ought at any rate to realize, that conformity with human nature can only be achieved at the expense of other desiderata.

BIBLIOGRAPHY

[Alvarez and Jaffee, 2005] L. Alvarez and K. Jaffee. Narcisissm guides mate-selection; human mate assortitavely, as revealed by facial resemblance, following an algorithm of 'like seeking like'. *Evolutionary Psychology*, 2: 177–194, 2005.

[Annas, 1993] J. Annas. Women and the quality of life: two norms or one? In M. Nussbaum and A. Sen (eds.), *The Quality of Life*. London, Oxford University Press, 1993.

[Ardrey, 1961] R. Ardrey. *African Genesis: A Personal Investigation into the Animal Origins and Nature of Man New York*, Macmillan, 1961.

[Axelrod, 1984] R. Axelrod. *The Evolution of Co-operation*. New York: Basic Books, 1984.

[Barash, 1979] D. Barash. *The Whispering Within, Evolution and the Origin of Human Nature*. Middlesex, Penguin, 1979.

[Barrett *et al.*, 1987] P. H. Barrett *et al.* (eds.). *Charles Darwin's Notebooks, 1836–1844*. Cambridge: Cambridge University Press, 1987.

[Baumeister, 2000] R. Baumeister. Gender differences in erotic plasticity: The female sex drive as socially flexible and responsive. *Psychological Bulletin*, 126: 347–374, 2000.

[Boehm, 2000] C. Boehm. The Origin of Morality as Social Control, in Katz, ed., pp. 79-101

[Boyd and Silk, 1997] R. Boyd and J. B. Silk. *How Humans Evolved*, New York, Norton, 1997.

[Buss, 1989] D. Buss. Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioural and Brain Sciences*, 12: 1–49, 1989.

[Cere, 2001] D. Cere. Courtship today: The view from academia. *The Public Interest*, 143: 53–75, 2001.

[Cronin, 1991] H. Cronin. *The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today*, Cambridge, Cambridge University Press, 1991.

[Crook, 1994] P. Crook. *Darwinism, War and History: The Debate over the Biology of War from the 'Origin of Species' to the First World War*, Cambridge, Cambridge University Press, 1994.

[Darwin, 1975] C. Darwin. *The Variation of Plants and Animals under Domestication*, 2 vols., 2nd ed., John Murray, London, 1975.

[Darwin, 1990] C. Darwin. *The Descent of Man*, London, Folio Society, 1990.

[Dawkins, 1976] R. Dawkins. *The Selfish Gene*, London: Granada, 1976.

[Dawkins, 1982] R. Dawkins. *The Extended Phenotype*, Oxford, Oxford University Press, 1982.

[De Waal, 1996] F. De Waal. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*, Cambridge, MA Harvard University Press, 1996.

[Dennett, 1996] D. Dennett. *Darwin's Dangerous Idea*, New York, Touchstone, 1996.
[Diamond, 1991] J. Diamond. *The Rise and Fall of the Third Chimpanzee*, Vintage, London, 1991.
[Fisher, 1992] H. Fisher. *The Anatomy of Love: A Natural History of Mating, Marriage and Why we Stray*, New York, Ballantine, 1992.
[Flack and de Waal, 2000] J. C. Flack and de F. B. M. Waal. 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. In Katz [2000], 1–29.
[Foucault, 1988] M. Foucault. *The History of Sexuality*, Vol. I, tr. Robert Hurley New York: Vintage, 1988.
[Frank, 1989] R. Frank. *Passions within Reason*, New York, Norton, 1989.
[Gigerenzer *et al.*, 1999] G. Gigerenzer, P. M. Todd, and the ABC Research Group. *Simple Heuristics that Make us Smart*, New York, Oxford University Press, 1999.
[Gould and Leuwontin, 1979] S. J. Gould and R. Leuwontin. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, Series B*, Vol. 205, No. 1161: 581–598, 1979.
[Hamilton, 1964] W. Hamilton. The genetical evolution of social behaviour I and II. *Journal of Theoretical Biology*, 7: 1–16, 17–52, 1964.
[Harris, 2004] C. Harris. The evolution of jealousy. *American Scientist*, 92: 62–71, 2004.
[Hrdy, 1981] S. B. Hrdy. *The Woman That Never Evolved*, Cambridge MA, Harvard University Press, 1981.
[Hrdy, 1999] S. B. Hrdy. *Mother Nature: Maternal Instincts and How they Shape the Human Species*, New York: Ballantine, 1999.
[Huxley, 1893] T. H. Huxley. *Evolution and Ethics*, London, Macmillan, 1893.
[Katz, 2000] L. D. Katz. *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives. Journal of Consciousness Studies*, 7(1 &2), repr. Imprint Academic, 2000.
[Keith, 1946] Sir A. Keith. *Evolution and Ethics*, G.P. Putnam's Sons, New York, 1946.
[Kellert, and Wilson, 1993] S. Kellert and E. O. Wilson, *The Biophilia Hypothesis*, Washington D.C., Shearwater, 1993.
[Kenrick *et al.*, 1990] D. T. Kenrick, E. K. Sadalla, G. Groth, and M. R. Trost. Evolution, traits, and the stages of human courtship: Quantifying the parental investment mode. *Journal of Personality*, 58: 97–116, 1990.
[Kitcher, 1985] P. Kitcher. *Vaulting Ambition: Sociobiology and the Quest for Human Nature*, Cambridge MA, MIT 1985.
[Maccoby and Jacklin, 1975] E. Maccoby and C. Jacklin. *The Psychology of Sex Differences*, London: Oxford University Press, 1975.
[Matthen and Ariew, 2002] M. Matthen and A. Ariew. Two ways of thinking about fitness and natural selection. *Journal of Philosophy*, 99, 55–83, 2002.
[Matthen and Linksy, 1988] M. Matthen and B. Linksy (eds.). *Philosophy and Biology*, Calgary, University of Calgary Press, 1988.
[Maynard Smith, 1982] J. Maynard Smith. *Evolution and the Theory of Games*, Cambridge, Cambridge University Press, 1982.
[Midgely, 1978] M. Midgely. *Beast and Man*, Hassocks, Sussex, 1978.
[Moore, 1903] G. E. Moore. *Principa Ethica*, London, Routlege and Kegan Paul, 1903.
[Moore, 1998] M. Moore. Nonverbal courtship patterns in women: Rejection signaling – An empirical investigation. *Semiotica*, 118: 201–214, 1998.
[Morris, 1967] D. Morris. *The Naked Ape. A Zoologist's Study of the Human Animal*, New York, Delta, 1967.
[Nicholas, 2005] L. H. Nicholas. *Cruel World: The Children of Europe in the Nazi Web*, New York, Borzoi Books, 2005.
[Penn and Potts, 1998] D. Penn and W. Potts. How do major histocompatibility complex genes influence odor and mating preferences? *Advances in Immunology*, 69: 411–36, 1998.
[Pinker, 2002] S. Pinker. *The Blank Slate: The Modern Denial of Human Nature*, New York, Viking, 2002.
[Ridley, 1994] M. Ridley. *The Red Queen: Sex and the Evolution of Human Nature*, Harmondswoth, Penguin, 1994.
[Rosenberg, 1991] A. Rosenberg. The biological justification of ethics: A best case scenario. *Social Policy and Philosophy*, 8: 86–101, 1991.
[Rosenberg and Sommers, 2003] A. Rosenberg and T. Sommers. Darwin's nihilistic idea: Evolution and the meaning of life. *Biology and Philosophy*, 18: 653–668, 2003.

[Ruse, 1986] M. Ruse. *Taking Darwin Seriously*, Oxford, Blackwell, 1986.
[Ruse, 1988] M. Ruse. Evolutionary ethics: Healthy prospect or last infirmity? In M. Matthen and B. Linksy (eds.), *Philosophy and Biology*, pages 27–74, 1988.
[Singer, 1982] P. Singer. Ethics and sociobiology. *Philosophy and Public Affairs*, 11: 40–64, 1982.
[Skyrms, 1996] B. Skyrms. *The Evolution of the Social Contract*, Cambridge: Cambridge University Press, 1996.
[Smuts, 1985] B. Smuts. *Sex and Friendship in Baboons*, New York: Aldine, 1985.
[Sober and Wilson, 1998] E. Sober and D. S. Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behaviour*, Cambridge, Harvard University Press, 1998.
[Trivers, 197q] R. Trivers. The Evolution of Reciprocal Altruism, *Quarterly Review of Biology* 46: 35–57, 1971.
[Trivers, 1972] R. Trivers. Parental investment and sexual selection. In B. Campbell (ed.), *Sexual Selection and the Descent of Man 1871-1971*, Chicago, Aldine, pages 136–179, 1972.
[Trotter, 1919] W. H. Trotter. *Instincts of the Herd in Peace and War*, London, Oxford University Press, 1919.
[Waddington, 1941] C. H. Waddington. The relations between science and ethics. *Nature*, 148: 270–74, 1941.
[Wedekind *et al.*, 1995] C. Wedekind, T. Seebeck, F. Bettens, and A. J. Paepke. MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London-Series B: Biological Sciences*, 260: 245–249, 1995.
[Wilson, 1975] E. O. Wilson. *Sociobiology*, Cambridge MA, Harvard University Press, 1975.
[Wilson, 1978] E. O. Wilson. *On Human Nature*, Cambridge MA, Harvard University Press, 1978.
[Wilson, 1986] E. O. Wilson. *Biophilia*, Cambridge MA, Harvard University Press, 1986.
[Wilson and Ruse, 1985] E. O. Wilson and M. Ruse. The evolution of ethics. *New Scientist*, 17 pp. 50–52, 1985.
[Wright, 1994] R. Wright. *The Moral Animal: Why We Are the Way We Are*, New York, Vintage, 1994.

Part III
Genetics

This page intentionally left blank

GENETIC ANALYSIS

Raphael Falk

1 INTRODUCTION

Genetic analysis is the art of analyzing the phenomena of heredity by hybridization. It has been introduced in 1865 by Gregor Mendel.

Plant and animal hybridization has been a common domestication practice since pre-historic times. In the late eighteenth century, following Linnaeus, hybridization became a major tool in scientific studies of taxonomy and the modifiability of species. Brünn (now Brno, in the Czech Republic) of the nineteenth century, a center of ambitious economic breeding that aspired to establish hybridization on sound scientific principles, exposed the young Mendel to both of its aspects [Orel, 1984; 1996]. Deeply involved in his ecclesiastical philosophy, as well in his scientific training and practical experience [Falk, 2001a], Mendel introduced in his talk to the monthly meeting of the Natural Science Section of the Agricultural Society of Brno [Mendel, 1866/1966] two new ingredients to hybridization work, which turned it into a powerful analytic research methodology:

1. Itemization of characters, viewing characteristics as *phenomena* per se, disregarding their particularities.

2. Analyzing the hybridization results as scientific data that should be described by *measurements* expressed in numerical terms.

These were the two elements that had been the core of the physical sciences for the last three centuries: Galileo observed the falling stone not as a stone but as a falling *body*, a phenomenon that could be generalized, and expressed his observations of this phenomenon in *numerical* relationships.

Although the discussions on the extent to which the science of heredity, established in 1900, was founded on Mendel's notion of a theory of heredity still rages [Falk, 1995b; Falk and Sarkar, 1991; Kottler, 1979; Meijer, 1985; Monaghan and Corcos, 1990; Olby, 1979; Stamhuis, 1995; Stamhuis *et al.* 1999; Theunissen, 1994], there can be little doubt that the *methodology of genetic analysis* was that introduced by Mendel in his paper of 1865. Genetic analysis has been dealing with the phenomenon of heredity by parsing it into unit-entities (traits or unit-characters — including DNA nucleotides — being *markers* of these entities), the inheritance of which could be followed by measurable numerical relationships

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

obtained through cross-breeding experiments. These principles applied when the Mendelian *Faktoren* — or *genes* [Johannsen, 1909] — were conceived as abstract intervening variables or as possible material-physiological hypothetical constructs [MacCorquodale and Meehl, 1948].

The chromosomal theory of heredity introduced by Thomas Hunt Morgan and his associates assumed that genes are discrete and distinct material unit-entities, and Herman J. Muller set out on genetic analytic research program to determine their physico-chemical properties (see, e.g., Muller [1922]). The nature of the hereditary material was, however, determined by chemical raher than by genetic analysis, and with the introduction of the Watson-Crick model of DNA structure [Watson and Crick, 1953a] genetic analysis became molecular. "Hybridization" was extended in prokaryotes (bacteria and viruses) beyond sexual mating, to include DNA transfer from organism to another either directly (transformation) or through a viral vector (transduction). With the advance of genetic engineering, sophisticated plasmids — small accessory bacterial DNA molecules, into which foreign DNA sequences can be inserted — were constructed. These extended the power of genetic analysis greatly, including the violation of the species-barrier of hybridization. Such plasmids function nowadays as vectors for the hybridization even of completely unrelated species of plants and animals. YACs (yeast artificial chromosomes) and BACs (bacterial artificial chromosomes) are the modern vectors of genetic analysis by molecular hybridization, both for the very sequencing of DNA-segments and for the elucidation of the functional aspects of such sequences. Furthermore, modern genetic analysis is not limited to "hybridization" *in vivo*. It has been extended to *in vitro* analyses, utilizing the annealing properties of homologous sequences of polynucleotides, like that of the PCR (polymerase chain reaction) method that amplifies specified DNA stretches (attached to primer sequences) by repeated cycles of dissociation of the DNA molecule and synthesis of the complimentary homologous sequence that reanneals with the dissociated one. Examination of homologies of nucleotide sequences by the hybridization of DNA and RNA strands thus maintains the hallmark of genetic analysis.

*

From its early steps, genetic analysis was paradigmatically reductionist. Although early genetic analysis was performed almost indiscriminately with any organism available, toward 1910 researchers converged on a relatively small number of species. By the 1930s the two major organisms for genetic and cytogenetic research of the mechanics of hereditary transmission were the fruit fly *Drosophila melanogaster* and maize (*Zea mays*). Experimental work was, of course also carried out with other organisms, mammals like mice, rats, guinea-pigs; insects like Chironomus and Sciara; and plants like Jimson weed (*Datura*), Snapdragon (*Antrirhinum*), peas, barley etc.

In 1936 George Beadle joined Boris Ephrussi in an effort to integrate embryology and genetics [Beadle and Ephrussi, 1936]. At the time, however, Drosophila turned out not to be an appropriate organism to follow the developmental physiological

or biochemical metabolic problems. Thus, Beadle joined Edward Tatum in the study of a simpler organism, the mold *Neurospora crassa*. Neurospora turned out to be not only extremely well adapted to the research of genetic control of metabolic pathways, it also significantly extended the power of genetic analysis of the mechanisms of inheritance.

In the 1940s Salvador E. Luria and Max Delbrück [Luria and Delbrück 1943] and Joshua Lederberg and Tatum [1946] went still further and showed that genetic analysis of bacteria was feasible. Thus, by the time Watson and Crick put forward their model of DNA structure [Watson and Crick, 1953b], genetic analysis of prokaryotes, mainly *Escherichia coli* and *Salmonella typhimurium* and their viruses (bacteriohages) was well advanced to accommodate the era of molecular genetics.

The peak of genetic research as a paradigmatic reductionist science, was attained with Crick's "Central Dogma" [Crick, 1958] of the role of genes as the one-way providers of information at the functional-molecular level, and with Dawkins' "selfish genes" [Dawkins, 1976] as the determinants of fitness in Darwinian evolution. However, turning the focus of molecular genetics back to eukaryotes in the 1960s soon revealed the inadequacy of the rductionist program of molecular biologists. As noted by Sydney Brenner, the paradigm of gene regulation "does not tell us how to make a mouse but only how to make a switch" [Brenner *et al.*, 1990]. The need to consider the impact of both ontogenetic and phylogenetic organizational constraints, increasingly forced molecular genetic analysis of the end of the twentieth century, to overcome the functional analyses of single genes for a top-down systems' analysis of 'developmental-biology', as well as to evolution.

Genes were no more discrete, unreducible entities [Beurton *et al.*, 2000]. And, besides the time-honored mouse and Drosophila, new organisms such as yeast (*Saccharomyces cerevisiae*), the nematode *Caenorhabditis elegans*, and the plant *Arabidopsis thaliana*, were added as experimental model organisms. Toward the end of the twentieth century and the Human Genome Project, humans too, or at least their DNA sequences, were added to the list of model organisms.

In what follows I wish to present the history of genetic analysis as the history of an internally consistent and continuously unfolding conceptual methodology of hybridization.

2 FROM FACTORS TO GENES: MENDEL TO JOHANNSEN

The nineteenth century, a period of great advances in the life-sciences, was inflicted with tensions between the introduction of reductionist notions and methodologies according to the model of the physical sciences, and the traditional vitalist and organismic beliefs in the existence of special laws of nature concerning living forms. Of the three great material theories of the life-sciences in the nineteenth century, Darwin's theory of evolution by natural selection, Schwann and Schleiden's cellular theory that culminated in Virchow's maxim *omni cellule e cellula*, and Mendel's theory of heredity by discrete factors, Mendel's was the most explicit reductionist

one. Furthermore, contrary to reductive explanations in, say, physiology, which were chemical and physical in nature, that offered by Mendel was purely phenomenological ("Developmental mechanics", although declared by Wilhelm Roux, its founder, to be the causal study of *form*, strived to reduce all phenomena to "the more recent concepts of physics and chemistry" [Roux, 1894/1986]).

Mendel's *Versuche über Pflanzen-Hybriden* (Experiments on plant hybrids) read at the February and March 1865 meetings of Brünn's *Naturforschenden Verein* presented a well formulated theory of heredity that he had probably conceived already before he carried out his hybridization experiments [Orel, 1996, 93–95]. The very small size of the patch of land at his disposal in the garden of his monastery (35m × 7m; plus some more space in Abbot Napp's greenhouse measuring 22.7m × 4.5m; see Orel, [1996, 96]) dictated him a meticulous planning of his experiments in which he obtained within seven years (1856–1863) results from many thousands of plants. There was absolutely no room for wasteful trial-and-error experiments. In 1936 R. A. Fisher pointed out the improbability of getting numerical results that fit expectations to the extent that Mendel presented in the published version of his talks (Fisher [1936]; see also Stern and Sherwood [1966]): His results were too good to be true. Hypotheses of foul play by Mendel or one of his helpers abounded (see, e.g., Sapp [1990, Chapter 5]). However, as already noted by L. C. Dunn, "the excessive goodness of fit to a theory that runs through his data certainly indicates that he had a theory in mind when the data as reported were tallied" [Dunn, 1965]. Recently Edward and Charles Novitski's analysis of Mendel's experimental results indicates that Mendel may have planned his experiments even more carefully than was believed so far: Anticipating the failure of some seeds to germinate or grow, Mendel, it appars, replaced failed plants by applying a of policy that inadvertently introduced a bias approximately compensating for the deviation that Fisher would have expected [C. E. Novitski, 2004; E. Novitski, 2004].

Claims that Mendel's objective was merely practical, namely to find the empirical laws that describe the formation of hybrids and the development of their offspring over several generations rather than uncovering universal laws of nature [Monaghan and Corcos, 1990] must also be rejected [Falk and Sarkar, 1991]. To the extent that one can reconstruct Mendel's intentions, I suggested that he was looking for the plan of the Creator for eternal transcendental realities. Like Kepler, two hundred and fifty years before him, he believed that this plan of reality is formulated in simple mathematical relationships [Falk, 2001a]. Unfortunately, because of unsuccessful experiments with *Hieracium* (the hawkweed) on the instigation of the botanist Nägeli, and his growing duties as abbot of the monastery, Mendel did not pursue his hybridization work [Orel, 1996].

Whereas the success of reductionism in physiological studies, induced embryologists to direct their efforts toward developmental mechanics (*Entwicklungmechanik*) [Lenoir, 1982], students of evolution could not agree on a simple reductionist mechanism that would replace Darwin's notion of continuous variation of the organism as a whole (note that Darwin's speculation of inheritance by particulate pangenes was a *pre*-cellular hypothesis. Pangenes migrated freely through the organism's

organs). Hugo de Vries's theory of *Intacellulare Pangenesis* [de Vries, 1889], developed in the last decades of the nineteenth century, which was both reductionist and preformationist, was intended to replace, or at least update Darwin's hypothesis of evolution by continuous variation. It had little impact when first published. However, at the turn of the century the attitude towards reductionism in the life sciences changed. When in 1900 de Vries "rediscovered" Mendel's paper he presented it as a significant contribution to his own conception of evolution and differentiation:

> According to pangenesis the total character of a plant is built up of distinct units. These so-called elements of the species, or its elementary characters, are conceived of as tied to bearers of matter, a special form of material bearer corresponding to each character. Like chemical molecules, these elements have no transitional stages between them. [de Vries, 1900/1966, 106].

De Vries explicitly suggested a bottom-up, reductionist alternative to Darwin's theory of evolution by natural selection. Whereas his program pointed to the process of species formation, Mendel's theory was for him merely a theory of inheritance [Brannigan, 1979]. Thus, he ignored Mendel's painstaking methodological design, like the selection of seven traits that could be gainfully analyzed by his experiments, and imputed to Mendel the confirmation of his viewpoint that "*the concept of species recede into the background in favor of the consideration of a species as a composite of independent factors*" [de Vries, 1900/1966, 108]. De Vries continued to promote his conceptions of evolution by discontinuous mutations — this time with considerable success — in the voluminous tome *Die Mutationstheorie: Versuche und Beobachtungen über die Entstehung von Arten im Pflanzenreich* [de Vries, 1902–3] (The English version was entitled *Species and Varieties: Their Origin by Mutation*; see de Vries [1912]). The title of volume I was an explicit rephrasing (in German) of Darwin's *Die Enstehung der Arten durch natüliche Zuchtwahl*, namely *Die Entstehung der Arten durch Mutation*, paying little attention to Mendel's hypothesis or experimental design.

On the other hand, Carl Correns — another "rediscoverer" of Mendel's overlooked paper — admitted that he had discovered for some time the same "regularity of the phenomena" as Mendel did, but "did not consider it necessary to establish [his] priority" [Correns, 1900/1966, 120], obviously did not appreciate the power of this analytic instrument. Thus, he joined, rather than established the science of heredity as a discipline of its own (see also [Falk, 1991; 1995b]). Although Correns contributed significantly to the experiments and theory of genetic analysis [Correns, 1924a; 1924b; Rheinberger, 2000b], his interest was primarily that of a physiologist, and rather less in the mechanics of hereditary transmission. (For the role of Tschermak in the "rediscovery" of Mendel, see Stern and Sherwood [1966, x-xii]; Olby [1985]).

The person who immediately grasped the analytic power of Mendel's work (upon reading de Vries' paper) for the study of *inheritance* was William Bateson. Bateson

at once shifted his research program to the study *genetics* — a term he introduced in 1905 — and aggressively promoted Mendel's conception of the rules of heredity and the methodology of hybridization. It would be no exaggeration to assert that he very much acted as Mendel's Bulldog, just as Huxley was Darwin's [Cock, 1973; Coleman, 1970; Darden, 1977; Falk, 1995b; Olby, 1987].

De Vries's confounding Mendel's methodology and his own ideology harassed the study of genetics at least during the first decade of the twentieth century, and arguably it still does today. Mendel did not employ in his explanations invisible particulate determinants, nor did he conceive of the difference between the potential for a trait and the trait proper. Instead, he studiously paid much attention to the pea varieties that would be adequate for his experimental design, noting that,

> Selection of the plant group for experiments of this kind must be made with the greatest possible care if one does not want to jeopardize all possibilities of success from the very outset.
>
> The experimental plants must necessarily
>
> 1. Possess constant differing traits.
> 2. Their hybrids must be protected from the influence of foreign pollen ...
> 3. There should be no marked disturbances in the fertility of the hybrids and their offspring in successive generations. ... [Mendel, 1866/1966, 3].

Of the 34 varieties of peas that he examined for two years he selected only twenty-two that "yielded quite similar and constant offspring" [Mendel, 1866/1966, 4]. Among these he picked out seven traits, each of which turned up regularly in one of two distinct alternative forms. In retrospect, for Mendel this procedure made the distinction between genotype and phenotype redundant [Falk, 1991].

De Vries' notion of the organism as a patchwork mosaic of discrete factors that keep their identity from one generation to the next, and segregate from each other in the gametes of hybrids [de Vries, 1900/1966] was enthusiastically adopted by Bateson. Like de Vries he contended the stepwise discontinuous evolutionary progress. According to Bateson's "Theory of Repetition of Parts" organisms are constructed as discontinuous repeats with variations of the same theme. Different organisms are systemic variations of such repeats [Bateson, 1894]. Together with his coworkers he examined the validity of the Mendelian principles of independent segregation of unit-characters for many traits in various plant and animal species as well as in humans. Unit-characters were defined as such by morphological considerations. He suggested that the two alternatives of unit character (which he called allelomorphs, later shortened to alleles) represented the respective presence and absence of the unit character [Schwartz, 2002]. However, upon encountering cases in which the structural or physiological unit characters did not correspond to the unit character of inheritance Bateson and coworkers resorted to *ad hoc* helping

hypotheses: To the extent that the morphologist's or physiologist's unit character did not agree with those of Mendelian segregation units, the former were not really "unit characters," but rather complex characters. Mendelian segregation became for Bateson a device to assess spurious morphological unit characteristics [Schwartz, 1998; 2002]. Such was the case of the "compound character" of birds' comb, which appeared in four "antagonistic" forms that were inherited according to Mendel's rule of two independently segregating unit-characters [Bateson, 1928].

Bateson suggested a resolution of the "outward" or zygotic morphological unit character, "being represented in the gametes by more than one factor" [Hurst, 1906]. When Bateson encountered cases of two or more unit-characters in which "the proportions do not accord with Mendel's assumption of random segregation," he interpreted these along his 1891 Theory of Repetition of Parts, namely that "reduplication or proliferation" of gametes may take place, and that "some factors are distributed according to one of the duplicated series and other factors according to the normal Mendelian system" [Bateson and Punnett, 1911].

A main concern for Bateson was to bolster the particulate theory of inheritance, as opposed to Karl Pearson's biometrician's belief in natural selection operating upon small continuous variations. His attempt to "Mendelize" was bound to come to a head-on collision with that of the logical positivist and author of the influential treatise *Grammar of Science*. Karl Pearson was drawn to the problems of inheritance by Francis Galton's *Natural Inheritance* [Provine, 1971]. The correlation coefficient between parents and offspring became for Galton and his followers the typical measure of the hereditary force. "A major epistemological characteristic of this approach was its purely descriptive character ... Karl Pearson's treatment of ancestral heredity was exemplary in this respect" [Gayon, 2000, 74-75]. Thus, although Bateson confronted Pearson with the theory of heredity, or genetics, concerning a discipline *within* the theory of evolution, Pearson's confrontations with Bateson may be reduced to a conflict on the grammar of science: Whereas Pearson developed statistical methods for the analysis of data on variation, presumably without being committed to any hypothetical speculation, Bateson's starting point was the hypothesis of particulate inheritance, or genetics.

In reality Pearson developed Galton's "stirp" theory of inheritance, according to which the traits of the individuals in one generation are the sum of the contribution of those of their ancestors, one half of each parent, a quarter of each grandparent, an eighth of each grand-grandparent, etc. On Mendel's hypothesis ancestors were helpful only to analyze the particulate status of an individual so as to allow the prediction of the offspring's properties of hybridization in generations to follow. As shown by Udny G. Yule already in 1902, analytically the two approaches converge, as long as one discerned *heredity* from *hybridization* [Yule, 1902]: Mendelism is concerned with hybridizations; Hybridization is the method to study the *hereditary* variation of specific characteristics between *individuals*. Heredity deals with the *population*-aspect of inheritance (Tabery [2004]; see also Section 7, below).

Early Mendelians' notions of the *Faktoren* were basically preformationist: They did not distinguish between the *Faktor* and the character proper. The parti-

cles were somehow representatives of the characteristics into which they develop (*Entwickeln* – unfold in German). Thus when A. K. Darbishire's hybridization experiments with mice in England [Darbishire, 1902], and William E. Castle's experiments with rats in the United States showed unexpected variation in fur color of the progeny of hybridization experiments [Castle, 1906], the discreteness of the Mendelian *Faktoren* was challenged. William Castle recounted:

> My own experimental studies of heredity, begun in 1902, early led me to observe characters which were unmistakably *changed* by crosses and so I have for many years advocated the view that the gametes are not pure in the sense expressed by Bateson. [Castle, 1919a, 126]

It was, however, the Danish pharmacist Wilhelm Johannsen who made the next critical step in conceptually setting apart traits, the *phenotype*, as the markers of *something* which is the inherent essence needed for traits, the *genotype* and its discrete entities, the *genes* [Johannsen, 1909]. Johannsen nourished a typologist conception. His problem was to settle the conflict between a normal or binomial distribution of the quantitative traits that he studied and the notion of a distinct, discreet Aristotelian type. Johannsen's answer was briefly that characters (the phenotype) vary continuously but heredity (the genotype) varies discontinuously [Roll-Hansen, 1978]. As a phenomenologist, but not a positivist, he came forward with hypotheses that allowed him to "analyze" [Provine, 1971, 95] Galton's and later his own quantitative data beyond the statistical presentation, and to suggest that Galton's partial regression to the mean was due to the initial mixture of essential (geno-)types in the population. In his bean experiments he easily picked out (after little inbreeding) such types, or "pure lines" in each of which regression to the mean of the type was complete. As a follower of Pearson's phenomenology Johannsen conceived of the genotype as the quantitative correspondent to the phenotype, being "the type beyond the average" [Falk, 2000b, 320]. The result of analytical experiments in genetics is "the upsetting of the transmission conception of heredity ... The personal qualities of any individual organism do not at all cause the qualities of its offspring" [Johannsen, 1911, 130-131]; The observable characteristics of individuals are the responses caused by the type which is inherited. Accordingly, Pearson was referring to phenotypes, whereas Bateson used the phenotype as the referent to the invisible genotype. The phenotypes of traits that segregated as Mendelian unit characters were "markers" or sign poles of the units of the genotype, namely of the *genes*. But by the term gene "merely the simple notion should be expressed that *something* in the gametes has or may have properties that condition or take part in the determination of a property of the developing organism. No hypothesis on the existence of this 'something' should therewith be construed or supported" [Johannsen, 1909, 124]. Johannsen never committed himself to the nature of these genes, whether they were *intervening variable* or *hypothetical constructs* [MacCorquodale and Meehl, 1948]; this, however, became a major issue in the years to come [Falk, 1986].

As noted, Mendel *selected* traits that display two qualitative different forms.

However, most traits of plants and animals display patterns of continuous variation. Did Mendel and the geneticists that followed him, inadvertently picked for their hybridization experiments those traits the factors of which segregate of each other, independently of the segregation of other factors, whereas the inheritance patterns of the continuously varying traits do not follow the Mendelian pattern of inheritance? The hybridization experiments of Nilsson-Ehle with oats and wheat varieties which differed in the color of their seeds indicated that what looked like a continuous "normal" curve of color-intensity distribution in the progeny of hybrids could actually be considered a binomial distribution ascribable to several discrete genes each of which endows similar phenotypes (color *versus* no-color). The color intensity of the seeds was dependent on the number of color-alleles of all genes involved. The continuous distribution patterns was ascribed to "environmental" effects that blurred to some extent the color determined by the number of alleles [Nilsson-Ehle, 1909].

It was, however, East who emphasized the *universality* of the *method* of genetic analysis, when he pointed out that as such it must be capable to handle quantitatively-varying characters as much as is did qualitatively-different traits:

> As I understand it is a concept pure and simple. One crosses various animals and plants and records the results. With the duplication of the experiments under comparatively constant environments these results recur in sufficient definiteness to justify the use of a notation in which theoretical genes located in the germ cells replace actual somatic characters found by experiments ... Mendelism is therefore just such a conceptual notation as is used in algebra or in chemistry ... [East, 1912, 633]
>
> Since qualitative characters were the ones that could be divided into definite categories they were the ones attacked. ... If Mendel's law is to be worth anything as a generality, therefore it must describe the inheritance of these characters ... [East, 1912, 637]

By the end of the first decade of the twentieth century the foundations of the science of genetics were well established. It exhibited not just a reductionist methodology for the analysis of heredity, but had explicit reductionist conceptual connotations, if not of organismic individuation as types, then at least of genes as the atoms of traits. Johannsen's conceptual differentiation of the phenotype and genotype allowed the study of the mechanics of heredity as well as the study of the developmental pathway from gene to trait. However, it must be admitted that although Johannsen emphasized that the phenotype was the integration of both genotype and the environment, his distinction of terms facilitated a more literal statement of Galton's Nature *versus* Nurture, which often established a dichotomy, rather than dialectic of the two. With the introduction of R. A. Fisher's Analysis of Variance, this biased juxtaposition of Nature *versus* Nurture, or of genotype *versus* environment, in which "interaction" was relegated to an annoying but unavoidable residuum, got a quantitative formulation. *Genocentricity*

or biased emphasis on the genetic component of characteristics became often a biological conception rather than an analytical tool.

The conceptual distinction between the potential for a trait — the genotype, and the trait proper — the phenotype, induced at least one scientist to overcome his reservations from the embryological preformationism that was inherent in the formal notion of genetic theory [Morgan, 1910a]. Thomas Hunt Morgan had been involved in the study of the causes of sex determination at the cellular level [Gilbert, 1978]. He saw in Johannsen's conception of disparate hereditary factors and traits a "developmental" way out of the de Vriesian notion of hereditary preformationism, allowing him to maintain an organismic approach to development while accepting a particulate theory of inheritance [Falk, 2003, 89]. For a while he valiantly attempted to uphold the unity of the contribution of hereditary factors to traits and the developmental processes in which the factors were involved. He tried to maintain a many-to-many perspective of factors and traits, but the need to refer (and identify) the factors through their phenotypes as "markers" demanded simplifications. So for example, Morgan tried to retain a genetic nomenclature that contained developmental notions. However, the more genes were discovered that were involved in a given characteristic, the more complex the nomenclature become. Eventually, he noted that the nomenclature oriented toward multiple developmental steps was "not sufficiently elastic to allow the introduction of new terms in the series" [Morgan, 1913a, 12]. Reluctantly, Morgan had to adopt the terminology that indicated an etiological relationship between genes and traits [Falk and Schwartz, 1993].

In compliance with Mendel's notion of uncovering the laws of heredity, irrespective of the specific trait involved, this terminology had far reaching consequences. In their *The Mechanism of Mendelian Heredity* Morgan and his students asserted that they "use cause here in the sense which science always used this expression, namely, to mean that a particular systems differs from another system only in one special factor" [Morgan *et al.*, 1915, 209]. Thus, by ignoring the detailed mechanical-causal notion of *Entwicklungsmechanik* of connections between causes and effects, Morgan and his students reduced causation to patterns conjunction. "The fact that correlations can be traced between the *end products* of ontogeny in successive generations ... is enough to declare that you have found the causes of these products" [Amundson, 2005]. This differential concept of the gene [Schwartz, 2000], which reduced the relationship to *changes* in genes and *changes* in traits, "has cut ontogenetic development out of the explanatory picture" [Amundson, 2005]. Morgan, Sturtevant, Bridges, Muller and many others, reduced the problems of development to the study of specific, differential factors of the ontogenetic reaction in a complex organic system, or to "genes for" specific traits [Falk, 1997].

3 THE CHROMOSOMAL THEORY OF HEREDITY

Already in 1896, in his *The Cell in Development and Inheritance* E. B. Wilson [1896] developed the theory of the cell as the fundamental unit of living organisms

and further interpreted August Weismann's hypotheses of the central role of the cellular nucleus, and especially its chromosomes, in development and differentiation [Allen, 1966, 1978; Griesemer, 2000]. Considering the dispute of embryologists between preformationists and epigenesists, Wilson concluded that "in some measure, a reconciliation between the extremes of both the rival theories" should be found;

> [W]e may consistently hold with Driesch that the prospective value of a cell may be a function of its location, and at the same time hold with Roux that the cell has, in some measure, an independent power of self-determination due to its inherent specific structure. [Wilson, 1893/1986, 78]

Since a correlation between sex and the pattern of the chromosomes, as they are revealed at cell division (mitosis), was found in many animal species, sex differentiation became a favorable subject for the elucidation of the role of chromosomes in heredity and development. In the cells of sexually reproducing species chromosomes may usually be arranged in a species-specific pattern of *pairs* — they are diploid. Following a sequence of two complex cells divisions with only one chromosome replication (meiosis) in gametogenesis, only one partner of each chromosome pair of somatic cell is found in the gametes. Thus upon fertilization of an egg by a sperm, the species-specific chromosome number and pattern, is restored. The main exception to this is that of the sex-chromosomes. Whereas as a rule, females carry a pair of X-chromosomes, in the males only one X-chromosome is present; the single X-chromosome has no partner or one that differs from it in size or form — the Y-chromosome (in some groups, like birds and Lepidoptera the femails are the ones that carry the single X-like chromosome).

The need for a full complement of chromosomes for normal development was demonstrated in experiments with fertilized eggs (zygotes) of sea urchin (*Paracentrotus lividus*) with excessive or missing numbers of chromosomes that developed abnormally [Boveri, 1902]. The individuality of chromosomes and their causal involvement in differentiation was indicated further by the specific abnormalities of progeny with specific excessive or missing chromosomes.[1] At the same time, Sutton, one of Wilson's students, showed that all chromosomes of the grasshopper *Brachystola magna* can be individually identified and that at meiosis the segregating chromosomes behaved as if each pair were alleles of a Mendelian factor [Sutton, 1903]. Sutton pointed out that if each pair segregated independently of the others, different combinations of the chromosomes (12 pairs in females) would produce a very large number of patterns ($2^{12}= 4096$), as expected of the Mendelian theory of variation.

[1]Twenty years later, Blakeslee [1922] succeeded in his hybridization experiments with the Jimson weed *Datura stramonium*, to differentiate the phenotype of the fruiting-bodies of all possible 12 trisomics — i.e., otherwise diploid plants with one of the chromosomes represented three times instead of twice — correlating each to the specific chromosome that was excessive.

Confirmation of the independent segregation of chromosome pairs was obtained in meioses of *B. magna* males that possessed, besides the single X-chromosome, another pair of chromosomes that could be individuated by the difference in their size: In 146 cells the smaller element of the unequal pair segregated to the same pole as the X-chromosome in meiotic anaphase, whereas in 154 the larger element segregated with the X-chromosome [Carothers, 1913].

Notwithstanding, all these studies did not prove the causal role of chromosomes in heredity and development. Chromosomes could be just another cell-morphology trait that segregates as Mendelian traits do. However, the observation of a white-eyed fruit fly, a *Drosophila melanogaster* variant whose progeny demonstrated in extensive hybridization experiments the "criss-cross" pattern of heredity (white eyed females mated to wild-type, red eyed males produced wild-type, red eyed daughters and white sons) as does the cytologically observed X-chromosome [Morgan, 1910b], initiated the intensive analysis of the mechanisms of inheritance of the Morgan School. Their working-hypothesis was that chromosomes play a causal function in the transmission and development of traits (except for the Y-chromosome that presumably was "empty" or inert — see section 6). Several other mutant stocks besides the *white* one were established (mutants are designated by italicized letters, w), each of which carried a mutant that was not allelic to *white*.[2] Yet they too were inherited in the "criss-cross" pattern, like the cytologically observed X-chromosome. This strongly supported the hypothesis that the cytologically observed maneuvers of chromosomes are the physical cause for the segregation pattern of these genes, and that the mutant traits of these stocks were due to mutations in different loci on the X-chromosome — loci of "sex-linked" genes. Other mutants of Drosophila could be grouped into two major non-sex-linked or "autosomal" linkage groups, presumably related respectively to each of the two pairs of larger chromosomes of the fly. Put differently, these findings pointed to the Chromosomal Theory of Inheritance: Chromosomes were the carriers of the genes, whatever the nature of these genes might be.

Calvin B. Bridges' paper that introduced the newly established journal *Genetics*, providing the experimental evidence for the chromosomal theory of inheritance, may be considered as the flagship of the methodology of genetic analyses [Bridges, 1916]. Rare exceptions to the "criss cross" inheritance of the characteristic sex-linked "markers" were observed: Instead of the expected "wild-type" daughters and vermillion sons from a cross of *vermilion* (v) females to wild-type males, an occasional vermillion daughter (or a wild-type son) was obtained. Crossing the primary exceptional vermillion daughter to regular wild-type males produced 3-4% secondary exceptional progeny, rather than the rare primary exceptions (primary exceptional males were sterile). Several females and males of the progeny continued in their turn to regularly produce secondary exceptional progeny (compared to the "criss-cross" expectations). Bridges hypothesized that the primary excep-

[2] As shown by "complementation tests", i.e., the compound heterozygotes a/b were phenotypically non-mutant or "wild type". Different mutants that do not complement are allelic and designated a_1/a_2 — see also section 5

tional flies were due to "chromosomal non-disjunction", that is to an occasional meiosis in an oocyte in which the two X-chromosomes moved to the same pole rather than segregating to opposite poles (producing XX and no-X eggs, instead of normal eggs with a single X). When such an oocyte was fertilized by a sperm carrying a Y-chromosome a primary exceptional matroclinous daughter, carrying XXY chromosomes was produced (and when a complimentary oocyte missing any sex chromosome was fertilized by an X-chromosome-carrying sperm was produced, a sterile primary exceptional patroclinous XO male was born — the "empty" Y-chromosome presumably was needed only for fertility, if a male was born). Bridges showed that his *genetic* analysis of the chromosomal patterns of the flies was unequivocally confirmed in *cytological* analysis of the chromosome patterns. Bridges analysis not only established the Chromosomal Theory of Mendelian Inheritance, but also provided a firm foundation for genetic analysis as the major tool for the phenomenological research of inheritance.

Bridges noted that if there was no preference of which possible pairs of the three sex-chromosomes in XXY females was paired at meiosis — and consequently was directed to segregate, the third unpaired chromosome always segregating at random — a maximum frequency of 33.3% secondary non-disjunction may be observed. The experimental results indicated that the two Xs pair (and segregate) preferentially to each of the Xs and the single Y. His experimental tools did not allow him to challenge this assumption. Almost thirty years later, in 1948 Kenneth Cooper showed that in stocks in which one of the two Xs is grossly rearranged (and therefore the two Xs can hardly pair at meiotic prophase), as much as 90% secondary non-disjunction may be obtained. This could be avoided when a one-armed Y-chromosome rather than the normal, two-armed Y chromosome was used. He hypothesized that a triad, of a Y-chromosome, with each X-chromosome paired to (and segregating from) another arm of the two armed Y-chromosome was produced, rather than two chromosomes pairing and the third moving at random. Such triads may give up to 100% secondary non-disjunction since both Xs are directed to segregate to the same pole, away from the Y-chromosome [Cooper, 1948], (see also Falk [1955]). Of course, no indications of the physical and/or chemical forces that cause specific pairing and segregation of the chromosomes were indicated. These were open to wide and often rather wild speculations.

Once the chromosomal theory of heredity was established, cytogenetics became an integral part of genetic analysis. Already in 1914 Muller discovered a fourth small linkage group in *Drosophila melanogaster*, as expected from the cytological observations of the flies, which own, besides the sex-chromosomes and two major autosomal chromosome pairs, also a pair of minute chromosomes [Muller, 1914]. Bridges suggested that the dominant sex-linked mutation that notched the margin of the wings in heterozygous females and that was lethal in males was actually a physical deficiency of a chromosome section [Bridges, 1917], although at the time it was below the resolution power of cytological techniques. In stocks that normally carried markers that segregate independently, dependent segregation of these markers in genetic experiments was indication for presumptive chromosomal

reciprocal translocations. Also *vice versa*, cytologically detectable major chromosomal aberrations — primarily reciprocal translocations — were confirmed by genetic analysis of deviation from the independent segregation of the markers in Drosophila and maize, especially after it was shown that X-rays were efficient means to induce aberrations (see section 4).

Aberrations became a major tool for the analysis of chromosomal mechanics. Thus, for example, it was shown that in triploid Drosophila flies — having three, instead of two sets of each chromosome — although all three chromosomes may be paired to each other in meiosis, only two of the three homologues pair at *each site* along the chromosome, and the pairing site most proximal to the centromere was the most significant in indicating the chromosomes that segregate to opposite poles [Anderson, 1925; Bridges and Anderson, 1925], supporting cytological conclusions that the centromere was the site that leads chromosome segregation. The availability of large chromosomal inversions or translocations that were shown cytologically to interfere with chromosome-pairing, and that were detected genetically to induce non-disjunction and recombination-suppression (see section 4), further established the need for chromosome pairing in meiosis in determining chromosome segregation [Dobzhansky, 1929; Dobzhansky, 1931].

Although some species like corn (*Zea mays*) had both relatively large chromosomes and numerous genes that were located on these chromosomes, most experimental species had either large chromosomes but few genes that were located on them (e.g., *Lilium sp.*), or abundant gene markers but chromosomes too small to discern any but the most conspicuous structural changes. The latter was the case with *Drosophila melanogaster*, the major research object of Morgan and his associates. This changed dramatically when it was shown that the 'spereme' — "the discoid structure of the chromosomes" of the Malppighian tubules (and the salivary glands) in Dipteran larvae [Kostoff, 1930], was in reality a discontinuous structure, the number of components of which corresponded to the haploid number of chromosomes [Heitz and Bauer, 1933; Painter, 1934], (see also Falk [2003, 106–107]). These turned out to be highly uncoiled, homologously paired polytenic chromosomes due to repeated endomitoses (without separation of the replicated polytenes). Thus providing extremely powerful cytogenetic magnification and resolution of the fly's chromosomes, since identity in the coiling outline of all replicated (and paired) strands gave a typical pattern of "bands" and "interbands" across the polytenous structures. This allowed the construction of detailed linear cytological maps of the chromosomes of *Drosophila melanogaster* [Bridges, 1935; 1938; Bridges and Bridges, 1939]. Maintaining the pairing of homologous "bands" in hybrids for chromosomal rearrangements forced typical chromosomal configurations, such as loops (for inversions), Ω-like structures (for deletions), or crosses (for translocations), etc., which made heterozygotes of chromosomal aberrations effective tools for superposing of the genetic and cytological maps. Further striking support for the chromosomal theory of heredity was the assignment of genes to specific "bands" of the polytenic chromosome map. For example, following a series of white-eyed stocks of Drosophila, which turned out to be small

deficiencies of up to several dozen "bands" of the polytenic chromosomes, allowed the assignment of one "band", 3C2, common to all deletions, as the site of the gene *white* [Demerec and Hoover, 1936].

Prokaryotes have no nucleus defined by a membrane, and no chromosomes that undergo concerted mitotic (or meiotic) segregation as do eukaryotic chromosomes at nuclear division. Notwithstanding, genetic analysis revealed the existence of chromosomal structures also for bacteria as well as viruses (for a time, they were called "genophores"). Processes in bacteria, similar to sexual mating or analogous to them, such as transmission of DNA fragments by vectors such as viruses or plasmids (transduction) or transfer of naked DNA fragments (transformation), or, in viruses, by mixed multiple virus infection of the same host cell, produced merozygotes — partial hybrids, or incomplete zygotes — which proved amenable to detailed genetic analysis (see the following sections). The genetically constructed chromosome map of bacteria was eventually visualized "cytologically" by radioactive labeling the multiplying chromosome [Cairns, 1963].

3.1 Genetic analysis of the molecular chromosome

The chromosomal theory of inheritance assigned genetic meaning to the cellular structures observed by cytologists. Although the chromosomes were microscopically observable only during cell division, cytologists have produced evidence that these were the highly condensed structures of permanent entities as expected by genetic analysis. With the advancement of molecular analysis the chromosomes were viewed as highly organized molecules of DNA, to which proteins are attached. Only toward the end of the twentieth century attention was directed at the roles of proteins not only in the organization of the chromosomes but also in their function.

Once molecular analysis of the chromosome became available molecular-signposts became important genetic-markers along the chromosomes. A long series of restriction enzymes each of which cuts the DNA strand in a specific pattern and at a specific sequence of 3-6 nucleotides provided the first DNA-level chromosome maps of Restriction-Fragments-Length-Polymorphism (RFLP). With the advancement of whole genome DNA sequences, single nucleotide polymorphisms (SNPs) as well as variability in the blocks of short repetitive DNA-sequences (micro-satellites) polymorphisms, became a most efficient mapping tool of genetic engineering.

Cytogeneticists looked for the functional genetic correlates of the chromosomes' morphological structures. Thus, the centromeres were identified the sites of the spindle-fiber-attachment at cell division, and the heterochromatically stained regions were recognized as regions of poor genetic contents, respectively. On the basis of genetic analysis considerations Muller defined another organelle at the tips of the chromosomes, the telomeres, that were specific monopolar end-structures, in difference from all other chromosome site, which were bipolar and "sticky" when broken [Muller, 1940; Muller and Herskowitz, 1954]. Repeated attempts were made to correlate the "banding" of the polytenic chromosomes with (individual) genes and the inter-band with inter-genes, as was later tried with the bands and inter-

bands of the stained vertebrate chromosomes (Q bands produced by quinacrine hydrochloride, G bands produced by Giemsa, etc.).

Sensitive physical methods applied to the chromosomes of wild-type and large-chromosomal aberration Drosophila stocks, established that eukaryotic chromosomes are composed of continuous DNA molecules running from one telomere, through the centromere, to the other telomere [Kavenoff and Zimm, 1973]. Such immense long molecules, it turned out, are replicated by multiple sites of replication-initiation distributed along the DNA molecule [Huberman and Riggs, 1968].

Molecular analysis identified the fundamental unit of the basic strand of the dense structure of the chromosome that is the relevant genetic structure. It has the same type of design in all eukaryotes, the DNA molecule ~ 200 base pairs of which are wrapped at intervals around octamers of histone-proteins as nucleosomes [Kornberg, 1974]. The density of these nucleosomes changes with chromosomal region and with the segment's activity.

Centromeres turn out to be specific DNA segments that are controlled by genes which ensure their properties of binding specific proteins for spindle attachment and the suppression of any "normal" genes activity. The telomeres, on the other hand, are regions of single-strand DNA that are maintained and replicated by a complex of RNA-telomerase enzymes. These enzymes systems do fulfill significant cell functions in controlling the length of the chromosome ends, that is essential for cell longevity, and hence for controlling both senescence and tumorogenesis [Blackburn, 1991; Cech, 2004].

4 GENETIC LINKAGE AS AN ANALYTIC STRATEGY

Notwithstanding the cytogenetic evidence for the chromosomal theory of inheritance, the major support for the theory was derived from the phenomenon of genetic linkage. Although linkage was not necessarily evidence for a chromosomal theory, or *vice versa*, the notions mutually supported each other.

Already early on, when researchers probed the extension of Mendel's rule of independent segregation of *Faktoren* in different species and for a variety of traits, it became obvious that often there were significant deviations from the expected independent segregation of traits. Was this due to some functional or developmental dependence of different traits on some common factor, or to an association of the factors themselves? In more modern terms: were these cases of developmental or physiological pleiotropy, in which the same factor is involved in more than one trait characteristic, or were Mendelian factors proper sometimes linked rather than always segregating independently of each other? Note that the very term "pleiotropy" (see Rieger *et al.* [1976]), like those of "penetrance" and "expressivity" (see Falk [2000a]; Sarkar, [1999]) invented later, was primarily a semantic technique to neutralize "disturbing" deviations from the dogma, even when eventually they may have served as handles to reach beyond the narrow essentialist and reductionist notion of basic Mendelism.

Carl Correns was probably the first to take notice of the phenomenon of the deviation from the rule of independent segregation of different traits, and to put forward a physiological explanation of differential lethality of gametes [Correns, 1902). But it were Bateson and Punnett who elaborated on the phenomenon and proposed a theory of "coupling" and "repulsion" — terms borrowed from electro-magnetic theories — of Mendelian factors along Bateson's 1891 "Theory of Repetition of Parts." Some gametes, but not others, were "plainly a consequence of some geometrically ordered series of divisions" [Bateson and Punnett, 1911]. Bateson, the most keen proponent of the particulate mechanistic theory of inheritance — see his opposition to Pearson's biometrical notions — adopted a non-mechanistic approach when confronted with deviations from independent segregation of these factors: As he did with the metameric phenomena he had been dealing with earlier, he interpreted the deviations from Mendel's rule in terms of field-theories, inspired by the electro-magnetic field-theories of Faraday and others that heralded the demise of the mechanistic paradigm and the resurgence of the organicistic conception (Coleman [1970], see Gilbert *et al.* [1996]) for the establishment of the biological field theory.

Contrary to Bateson, Punnett and their colleagues' complete oblivion to cytological observations, Morgan found that "[t]he discovery that there occurs in the formation of the germ-cells a process that supplies the machinery by means of which segregation might take place has aroused ... interest in the application of the observations of cytology to the conclusions in regard to Mendelian segregation" [Morgan, 1910a, 465]. For Morgan the Chromosomal Theory of Inheritance provided the resolution to the phenomenon of deviation from independent segregation of trait, or linkage [Morgan, 1911].

> I venture to suggest a comparatively simple explanation based on results of inheritance of eye color, body color, wing mutations and the sex factor for femaleness in *Drosophila*. If the materials that represent these factors are contained in the chromosomes, and if those factors that "couple" be near together in a linear series, then when the parental pairs (in the heterozygote) conjugate like regions will stand opposed. ... [W]hen the chromosomes separate (split) ... the original material will, for short distances, be more likely to fall on the same side of the split, while remoter regions will be as likely to fall on the same side as the last, as on the opposite side. ...
>
> [W]e find coupling in certain characters, and little or no evidence at all of coupling in other characters; the difference depending on the *linear distance* apart of the chromosomal materials that represent the factors. ... The results are a simple mechanical result of the location of the materials in the chromosomes, ... and the proportions that result are not so much the expression of a numerical system as of the *relative location* of the factors in the chromosomes. [Morgan, 1911]

For the chromosomal theory of inheritance to be the explanation of linkage,

it was necessary to spell out why linkage was not absolute, or, why all factors linked to a given chromosome should not segregate as a unit. Since "coupling" and "repulsion" were not absolute, Morgan had to add the notion of crossing over, which he borrowed from Janssens' chiasmatype theory, based on observations of meiosis at spermatogenesis in the Batarchine salamander [Janssens, 1909]. During meiosis homologous chromosomes (each had already replicated to form two chromatids) pair, and are increasingly coiled and twisted one about the other up to the pachytene stage. At the diplotene stage the tension appears to be relieved and the paired chromosomes start to segregate, except at some sites, at which sister chromatids appear to cross over. Janssens hypothesized that the mechanical twisting leads to breaks in the paired chromosomes, allowing some relaxation of the tension by local untwisting, which is followed by repair of the broken "sticky" ends, at least some of such repair occurs between the paired-chromosomes. This mechanism of exchange provided for the partial linkage that Morgan had to explain.

> We may make a general statement or hypothesis that covers cases like these, and in fact all cases where linkage occurs: viz. that when factors lie in different chromosomes they freely assort and give the Mendelian expectation; but when factors lie in the same chromosome, they may be said to be linked and they give departures from the Mendelian ratios. The extent to which they depart from expectation will vary with different factors. I have suggested that the departures may be interpreted as the distance between the factors in question. [Morgan, 1913b, 92-93]

Thus, when Sturtevant harnessed genetic analytic experimentation to construct linkage-maps, he had in mind the mapping of genes along chromosomes [Sturtevant, 1913]. But these were actually virtual maps, the physical meaning of which was not at all generally accepted. Richard Goldschmidt pointed out that describing linkage data in terms of a linear map is merely the application of a standard procedure of analytic geometry. He doubted the inspiration based on the cytological observations of Janssens and suggested instead a physiological mechanism in which an "immune-like" attraction between chromosomes may occur at replication, the strength of linkage expressing the intensity of the reaction rather than the physical distances between sites [Goldschmidt, 1917]. Muller, however, succeeded in examining recombination in specially constructed multiply-linked genetic marked Drosophila stocks. He demonstrated that "factors behave as though they are joined in a chain; when interchange takes place, the factors stick together in sections according to their place in line and are not interchanged singly". Accordingly linkage is a phenomenon of the topology of the genes rather than a property of individual genes [Muller, 1916, 366]. Regarding William Castle's doubts "[t]hat the arrangement of the genes within a linkage system is strictly linear seems for a variety of reasons doubtful", for example, it was doubtful "whether an elaborate organic molecule ever has a simple string-like form" [Castle, 1919b, 26], Muller pointed out that the genetic analysis of linkage was a phenomenological study.

> Whether or not we regard the factors as lying in an actual material thread, it must on the basis of these findings be admitted that the forces holding them linked together — be they physical, 'dynamic' or transcendental — are of such nature that each factor is directly bound, in segregation, with only two others — so that the whole group, dynamically considered, is a chain. ... no implication as to the physical arrangement of the genes is intended when the terms 'linear series,' 'distance,' etc., are used. [Muller, 1920, 101]

Muller's statement notwithstanding, the correlation of the virtual linkage map and the physical chromosomal map was further supported by studies like that of Mohr [1924] on the *Notch*-"deficiency". As noted (in section 3), Bridges [1917] suggested that the dominant sex-linked, recessive lethal mutation *Notch* was actually a physical deficiency of a chromosome section. Notched females mated to white-eye males or to facet-eye males (*w* and *fa* are two recessive mutations at different but closely linked genes) produced notched-white and notched-facet females respectively. Mohr showed that the frequency of crossing over between genes lying to the left and to the right of the "deficiency" was significantly reduced in its presence. Once polytenic chromosomes of Drosophila (and other dipterans) were mapped, it was straight forward to correlate the genetically observed changes of the virtual linkage maps with cytological events. Thus, the genetic marker Notch turned out to be a deficiency of band 3C7. And the well known *C*-factors, or suppressors of recombination of genetic analysis (like in the *ClB* stock of Muller) turned out to be inversions of at the cytological level (see Lindsley and Grell [1968]). Most intra-inversion crossing-over events in hybrids of an inverted and a non-inverted chromosome produce upon recombination aneuploid, non-viable deficient or duplicated chromosomal products, and are consequently efficient genetic "suppressors" of recombination.

According to Janssens' mechanistic conception that was adopted by Morgan, linkage was primarily a function of the probability of breakage of the twisted chromosomes; the further apart genes were positioned on the chromosome the higher the probability of a random breakage and recombination event, and the looser the linkage (when it turned out that the frequencies of genetic recombination were not homogeneously distributed over all cytological unit length, this was modified, claiming that the structure of the chromosomes was not uniform, so that there were sections that broke more readily than others). Bridges [1916] analysis of non-disjunction indicated that exchange involved only two of the four strands at every recombination site. This was confirmed by L. V. Morgan who showed that in a Drosophila stock in which the two X-chromosomes were physically attached (and accordingly all their daughters were obligatory X-chromosome non-disjunctional, i.e., carried two of the four products — half a tetrad — from the *same* meiotic event), some daughters were homozygous for genetic markers for which the mother was heterozygous. This could be obtained only if cross-over occurred after chromosome replication between two of the four paired chromatids, when a non-crossover and a cross-over chromatid happened to disjoin. Furthermore, since according to

the model, the frequency of such marker-homozygotyzation was proportional to their distance from the centromere, it allowed the mapping of the centromere — the region of the chromosome with which the spindle fibers become associated, i.e., a cytological entity — on the virtual linkage maps [L. V. Morgan, 1922]. No information on the mechanism was available except that recombination could occur between any pair of the four chromatids, and that the closer the markers were the more one exchange interfered with another one nearby, which indicated rigor of the pairing chromosomes as expected in a mechanical model. These conclusions were upheld and extended when recombination in fungi, primarily in *Neurospora crassa*, were performed. In ascomycete fungi each meiotic division is confined to a discrete sac, or *ascus*. Micro-dissection of asci allows full "tetrad analysis," following all the products of individual meiotic events (contrary to the situation in most other organisms where the mass of all products of many meiotic events are observed, except for special cases, like attached-X, where "half tetrad-analysis" is possible). Furthermore, in some species, like Neurospora, the linear arrangement of the ascospores within the ascus allows discerning the products of the first meiotic division from those of the second meiotic division (see Fincham *et al.* [1991]).

The observation that both chromatids of the paired chromosomes could be involved in crossing-over events along the chromosome appeared to refute alternative to the mechanistic explanation, such as that exchange occurred between the newly produced chromatids, due to "copy-choice" at replication [Belling, 1928]. Additional notions, like that of mutual induction of sites on the chromosomes, or "gene conversion" [Winkler, 1930] were excluded by Barbara McClintock in maize [Creighton and McClintock, 1931] and Curt Stern in Drosophila [Stern, 1931], when they constructed stocks that allowed to follow both cytological physical exchange and genetic recombination of markers in the same individuals (see Falk, [1995a; 2003]).

Linear linkage maps of distinct genes entities, among which recombination may take place, became the dogma of genetic analysis. However, the resolution power of recombination analysis was limited by the shear size of the experiments that could be carried out: as a rule, few scores of progeny in mouse recombination experiments, and not more than some hundreds in Drosophila or maize. This changed, however, in the 1940s when fungi — primarily *Neurospora crassa* — and later bacteria and their viruses became available for genetic analysis. Screening methods were designed in which only (or almost only) the expected class of progeny, such as recombinants between given markers, survived.

For many years it was doubted whether bacteria, who do not have a distinct nucleus (they are prokaryotes), who multiply by fission, rather than by ordered mitosis-like chromosome segregation, and in which adaptive (apparently "Lamarckian") heritable changes may be induced by changing culture conditions, obey the rules of genetics as do nucleated organisms (eukaryotes). The evidence brought by Luria and Delbrück [1943] that heritable changes in bacteria populations were due to preadaptive mutations (i.e., "Darwinian" selection of preexisting mutations, rather than "Lamarckian" adaptive responses to environmental challenges),

and by Lederberg [Lederberg, 1947; Lederberg and Tatum, 1946] for a process of sexual (or parasexual) mating and recombination of genetic factors in the K-12 strain of *Escherichia coli*, convinced geneticists that bacteria might be amenable to "classical" genetic analysis along similar lines as have been carried out in eukaryotes. The feasibility of genetic analyses, such as the construction of linkage maps, was taken as sufficient evidence for chromosome-like existence also in prokaryotes [Lederberg, 1947; Lederberg *et al.*, 1951].

Two presumably unrelated observations eventually challenged the mechanistic notion of recombination: a. When recombination between very close markers was examined, it turned out that by far too many double recombination events occurred; instead of the expected increase of positive interference with decreasing distance, in very short intervals, *negative* interference was observed. b. When products of individual meiotic events could be followed, as became possible with "tetrad analysis" in fungi, infrequent but consistent deviation from the expected 1:1 recovery of the two alleles for which the parent was heterozygous turned up. This phenomenon of "gene conversion" — for a long time suspected to be an experimental "contamination" — was found to be often associated with recombination of outside markers. Thus, instead of finding that a heterozygote A/a produced 4 A spores and 4 a spores in the ascus (spore-sac, containing the products of a single meiotic division, followed in many species by a mitotic division) of *Neurospora*, some asci contained $6A{:}2a$ ascospores, or even $5A{:}3a$ ascospores, whereas the nearby outer markers (which often were found to recombine) segregated normally, giving $4A{:}4a$ ascospores [Lindegren, 1953; Mitchell, 1955].

The results with recombination in bacteria "illustrated the experimental concordance of bacterial segregations to a generalized definition of mendelism," although "[i]n a purely formalistic way, these data could be represented in terms of a 4-armed linkage group." Formally a branched linkage map emerged, however, Lederberg *et al.* added: "without supposing for a moment that this must represent the physical situation" [Lederberg *et al.*, 1951, 416-417]. Yet, consistent non-branched linear linkage maps were constructed once linkage maps were derived by measuring the *time of transfer* of marker genes in mating different Hfr (high frequency recombination) "male" bacteria to an F^- "female" bacterial strains. The virtual maps derived from different Hfr stocks proved to differ merely in the permutations of the same linkage relationships. It was concluded from the genetic analysis of such matings that the bacterial chromosome was a closed circle, which was "opened" at different sites in different Hfr stocks [Hayes, 1964; Jacob and Wollman, 1961]. This was eventually confirmed by cytological observations of the physical organization of the DNA of *E. coli* [Cairns, 1963].

The mechanical chiasmatype notion of recombination assumed that exchange occurred precisely at the same site in the two chromatids involved, since as a rule both products of recombination were viable. This, however, was not strictly necessary at the DNA molecular level. It was Whitehouse who first suggested a model of crossing-over, still based on breakage and rejoining of chromatids, in which the breakage points in the two strands of each of the exchanging DNA molecules were

staggered. Whitehouse suggested that the rejoining of the DNA molecules took place not by an end-to-end association of broken chains but by lateral association of complementary segments from homologous regions, to give — when exchange of chromatids took place — short *hybrid DNA* sequences [Whitehouse, 1965, 318]. Holliday [1964] proposed a different model, according to which the initial severance in DNA was enzyme-induced in one chain of the same polarity in each of the chromatids involved. According to this and later more sophisticated models, heteroduplexes — discordances of one or more nucleotide bases in paired strands of the DNA — could occur in such hybrid-DNA segment, which calls for the involvement of DNA-repair mechanisms. Such repair, when in the "right" direction, would reconstitute the 2:2 relationship between the products of this meiosis. However, in the "wrong" direction it would end up in a 3:1 segregation, i.e., in gene-conversion (suppose that an A-T nucleotide base pair in one chromatid is confronted by a G-C pair in the other, in otherwise identical sequences in both chromatids. If the hybrid DNA segment includes the above mentioned discordance, mispaired A-C and G-T heteroduplexes would occur. If one heteroduplex is repaired to become A-T and the other to become G-C, the 2:2 segregation at that site will be reconstituted. If, however, both heteroduplexes are repaired to become A-T, a 3:1 segregation ensues). Thus gene conversion was interpreted as a misrepair in an interaction between chromatids at meiosis. This interaction between two chromatids, involving a hybrid DNA segment formation (with or without heteroduplexes) may be resolved as an exchange between chromatids, or no-exchange. Once these notions were introduced, the mechanical breakage-rejoining model of recombination was replaced by recombination as an enzyme-driven process, amenable to biochemical analysis (see Stahl [1979]).

The insight gained from "the novel feature", namely the property of single chains of nucleotides to pair by the matching of purine (adenine – A, guanine – G) and pyrimidine (thymine – T – replaced by U – uridine in RNA, cytosine – C) bases, DNA with DNA or RNA, was now extended to become not only a powerful analytic tool of the mechanics of the chromosomal phenomena of heredity, but also a major instrument of genetic engineering. Genetic analysis and its application has become the science of hybridization at the molecular level.

5 GENES — THE ATOMS OF HEREDITY: FROM MULLER TO WATSON AND CRICK

In 1956, in the Brookhaven Symposium, Herman J. Muller succinctly put up "two basic questions, both as old as genetics but still warmly disputed".

> First, have we the right to speak of individual genes as separate bodies rather than only as convenient mental isolates, conceptually cut out of an uninterrupted genetic material, or chromosome, of dimensions far larger than they? Second, if the expression "the individual gene" does correspond to a material reality ..., do we have a right to regard any,

> or many, of the Mendelian differences with which we deal in genetics as representing changes, chemical in nature, within these individual genes; or are these differences caused only by decreases and increases in the number and possibly in size ..., and by changes in mutual arrangement of genes that themselves remain essentially unchanged, or even unchangeable? [Muller, 1956, 126]

Once Johannsen established the distinction of phenotype and genotype, it was manifest that genetic analysis would strive to deduce the nature of the genotype from observing the phenotype. The chromosomal theory of inheritance provided a material basis for Mendel's theory, and indeed, the theory of linkage suggested the mapping of these bodies. The early virtual maps became very real after the discovery of the polyteneic chromosomes of Drosophila, and genes, defined by their (phenotypic) effects, were reference points on these maps. Most geneticist treated genes as instrumental entities, without committing themselves to the nature of these entities: They were factors on the chromosomes amenable to genetic analysis. But in 1922 Hermann J. Muller in a programmatic paper [Muller, 1922], laid out his notion of the genes as the atoms of heredity, as well as the strategies that should be adopted to elucidate their structure and function. For Muller genes were discrete and distinct entities of matter. He explicated their three essential properties: a) self-replication, or *autocatalysis*, b) mediating functions and constructions of living matter, or *heterocatalysis*, and c) capacity of undergoing rare changes without loosing their property of self replication, or *mutability*. Whereas the first two properties were those immanent to hereditary factors per se, the third one was essential for organisms that evolve by a Darwinian process of variation and selection. Although he looked forward for the day when the gene would be "grounded in a mortar", that is, analyzed by physico-chemical methods, at the present time the only way to learn of their structure and function was by the methods of genetic analysis. Consequently he set out on a research program focused on the third property of the genes, which was unique to entities of heredity in a Darwinian system, namely the study of mutations. Following Troland's [1917] suggestion, Muller believed for some time that genes were enzymes, or at least acted like enzymes.

A diametric opposite conception of the genes was that of Richard Goldschmidt, who maintained a holistic, organismic, top-down conception. According to him there were no materially discrete and separable physical genes along the chromosomes. The chromosomes were integral units of function in a hierarchy of fields of genetic action, along which one may identify centers of specific functions that if changed (deleted or rearranged) affected a specific function of the whole. The local injuries of the chromosomes appeared as if they were discrete genes (Goldschmidt [1938; 1954; 1955], see also Dietrich [2000]). Although Muller claimed early on that "no implication as to the physical arrangement of the genes is intended when the terms 'linear series', 'distance', etc., are used" [Muller, 1920, 101], he himself repeatedly used physical and chemical considerations, arguing that "a little consideration, however, showed", that "direct chemical influences of the type in

question could hardly extended over distances so relatively vast from the chemical standpoint, as are those here in question" [Muller, 1938, 587].

It is important to remember that the basic empirical test for differentiating genes as distinct units was a *functional* test, the "complementation test": Two (recessive) mutations were confronted with each other in the same cell in heterozygotes (or heterokaryons in fungi, merozygotes in bacteria or multi-infection in viruses). If the phenotype of the compound heterozygote was normal ("wild type"), the mutations affected different genes; if, however, the compound heterozygote was phenotypically mutant, both mutations affected in the same gene (even if their own phenotypes were different). The *structural* properties derived from such a defined entity were the ones disputed. One piece of evidence against mutations being all deletions or rearrangements of more integrative entity, as claimed by Goldschmidt, was the observation that the same gene could mutate to several alleles, each affecting the same characteristic but at different intensity, like mutants of the gene *white* the alleles of which varied for eye color from white to apricot, buff, eosin, coffee, coral, ivory, blood, etc. (denoted $w, w^a, w^{bf}, w^e, w^{cf}, w^{co}, w^i, w^{bl}$, respectively). This, however, could be settled if one adopted Lewis Stadler's [1954] *operational* model of the gene as a linear series of points within a continuum, rather than an indivisible atom of inheritance, so that affecting (deleting) any one may cause functional disturbances, which may be observed as mutations of the organism. Muller found further support for the "atomic" gene concept in "position effect", where it was shown that not a deletion of the gene but, at least in specific cases, its rearrangement with respect to its neighbors, was responsible for the mutated effect [Raffel and Muller, 1940]. According to him, these "apparent mutations of genes located near points of chromosome breakage are only changes in gene functioning conditioned by an alternations of gene grouping" [Muller, 1938, 587]: Changing function by changing neighbors, has meaning only if neighbors do exist. Goldschmidt, on the other hand, claimed position-effects to provide strong evidence against the discreteness of the genes; it supports functional integrity rather than structural discreteness. Rearrangements that resulted in position-effect partial inactivation of genes (such as w^{M4}, white-mottled eye color) yielded upon X-ray irradiation new alleles of various grades of inactivation or reactivation, depending on the amount of heterochromatin (condensed chromosome segments, of poor genetic activity; contrary to euchromatic segments) present next to the site of the original gene [Panshin, 1935; 1936]. Muller considered this as evidence for the change in the configuration of genes due to the variation in their milieu that resulted in changed gene activity, analogous to that known for enzyme activity. However, Muller's argument for genes that may mutate (and not just be deleted) was an *a priori* one, anchored in his Darwinian notion of evolution by small discrete changes, according to which "it is not inheritance and variation which bring about evolution, but the inheritance of variation, and this in turn is due to the general principle of gene construction" [Muller, 1922, 35]. Genes, not whole chromosomes must be the units of evolution. He found further evidence for discrete genes being the units of evolution in the fact that "wild type" alleles were as a rule

dominant to most mutated alleles of known genes which were at least partially recessive. "Wild type" alleles were the ones selected to allow a buffered, more or less constant activity even in the presence of disturbances in the internal or external circumstances: "The degree of dominance [of the wild type allele] was probably only a concomitant or reflection of the general degree of activity ('potency') of the gene" and "lies in the protection thus afforded the individual against disturbing genic and exogenic influences" (Muller [1935, 409-410], see also Muller [1950]).

However, to study the nature of the process of mutation proper, a quantitative method was needed to measure mutation frequencies as function of varying a controllable agent. Attempts to learn about the nature of mutations by artificially inducing heat (and cold) shocks and similar treatments failed, because if they had any mutagenic effects these were too weak to be amenable to quantitative genetic analyses. Muller chose lethals as a *class* of mutants that abolish essential functions of genes (on the X-chromosome of Drosophila), thus providing an easy, unequivocal test for detecting mutations in any of many genes. The *ClB* chromosome was a specially designed trap for detecting lethal mutations in the fly's X-chromosome that allowed for the quantitative analysis of mutation induction [Muller, 1927b, 1928]. In 1927 Muller announced at the International Congress of Genetics in Berlin that he induced mutations by irradiating the gonads of Drosophila with intensive doses of X-rays [Muller, 1927a]. In the same year, Stadler showed that ionizing irradiations of barley pollen induced mutations in progeny plants [Stadler, 1928].

The quantitative analysis indicated that gene mutations were induced proportionally to the X-ray dose given to the spermatozoa of Drosophila flies, independent of the intensity (dose/time) of the irradiation, and apparently without a threshold dose. The induction of chromosomal aberration, on the other hand, was related exponentially to the irradiation dose. Thus it was inferred that mutations were single-hit events or "point mutations"; whereas aberrations required two independent single hit breaks (although the exponent was more like 3/2 than the expected 2). These conclusions were further supported by the observation that more "dense" ionization radiations, like those of neutrons, were highly efficient in inducing aberrations.

The "target theory" successfully determined the dimensions of bacteria and enzyme molecules from X-ray induced inactivation curves. The single-hit lethal-mutations' induction curve suggested the extension of the "target theory" to measure the dimensions of genes as if they were autonomous discrete (globular) entities in the cellular broth [Timoféeff-Ressovsky *et al.*, 1935]. This conception was reflected in Schrödinger's [1962] description of genes as "aperiodic crystals" and mutations as "quantum jumps" of states of matter in his influential *What Is Life?*

These estimates were, however, soon challenged when it was found that factors like anoxia reduced the frequency of mutations and chromosome breaks induced by a given dose of radiation. This meant that the radiation induced ionizations might induce mutations or chromosome breaks not necessarily by direct hit of the "target" itself. The effective target of the radiation could be larger than the final gene-

or chromosomal-event. If, however, the notions expressed in the "target theory" paper (known sometimes as the TZD experiment, in covert analogy to the EPR experiment by Einstein Podolsky and Rosen in physics) had the claimed impact on the physicists community to turn to biology for laws of nature not detected in physics research [Stent, 1968], then its bearing on the upcoming molecular biology age is profound indeed.

Although the research program to deduce the physico-chemical nature of genes from the genetic analysis was extended to include the study of chemically induced mutations, it turned out that this too did not resolve the nature of the genetic material (see section 5.1 below).

As Muller anticipated, many problems, like that of the nature of induce mutations, disappeared with the molecular genetic analysis. Once the chromosomes were recognized as continuous strand of nucleic-acids, the extended dispute of whether induced mutations were genuine "point mutations" became immaterial: In such a molecule there is no difference between, say, a deletion of one nucleotide or its conversion into another nucleotide.

Morgan's emphasis in his Nobel Lecture that for all practical purposes it did not make any difference what the physico-chemical nature of the genes were, may be taken as a subtle rebuke of Muller's research program:

> Now that we locate [the genes] in the chromosomes are we justified in regarding them as material units; as chemical bodies of a higher order than molecules? Frankly, these are questions with which the working geneticist has not much concerned himself, except now and then to speculate as to the nature of postulated elements. ... at the level at which the genetic experiments lie, it does not make the slightest difference whether the gene is a hypothetical unit, or whether the gene is a material particle. [Morgan, 1935]

Nevertheless, in spite of the failure to determine by genetic analysis the nature of the gene, Muller's focus on this problem widely opened the path of genetic analysis as a method for resolving the role of inheritance in organisms' function and development, as well as in evolution (see, e.g., Falk [1997]), in an effort to understand "the gene as the basis of life" [Muller, 1929; 1947]. In that sense he laid the foundations for the molecular age of genetic analysis (see Lederberg [1991]). Yet, the problem that Muller put up, as cited at the beginning of section 5, whether individual genes are separate bodies or only convenient mental isolates, prevails and obtains new significance at the age of genomics [Muller, 1929; 1947], (see Beurton *et al.* [2000]; Burian [2004]; Falk [2004]; Moss [2003]).

Beadle and Tatum's notion of a one-to-one relationship between genes and enzymes, the "one gene — one enzyme" slogan [Beadle and Tatum, 1941a, 1941b], was one direct consequence of the notion of genes as distinct entities along the chromosome. So was the notion of population genetics, the mathematics models of which, starting with the Hardy–Weinberg's rule of 1908, related to genes as the indivisible atoms of heredity (see section 8).

Toward the 1950s, with the improvement of the resolution power of genetic recombination experiments (see section 4), the internal linkage relationships of the genes themselves started to unfold. Altenburg and Muller [1920] called attention to a complex gene of Drosophila, *Truncate*, being a variable character, both somatically and genetically [Carlson, 1959]. Its complex nature, exhibiting combinations of three presumably independent phenotypes, was not resolved at that time. E. B. Lewis conceived of closely linked but recombinable genes of Drosophila, such as *Stubble-stubloid, Star-asteroid* or *bithorax-Ultrabithorax-bithoraxoid,* as "nests" of genes of similar function, possibly due to ancient duplication of one gene followed by its functional diversification with evolutionary time. He tried to extend this notion, when recombinants were detected in apparently non-complex loci such as *white* and *lozenge* Drosophila eye-color genes. Rarely recombination occurred between two mutants that did not complement each other —hence by definition, being of the "same gene" — when juxtaposed in a compound heterozygote (in *trans*: a_1 +/ + a_2). The phenotype of the recombinant in which the two mutations were now on the same homologue chromosome (in *cis*: a_1 a_2/ + +, the other homologue chromosome being non-mutant), was wild type. Lewis called these mutants "pseudo-alleles.". He suggested that pseudo-alleles belonged to distinct adjacent genes, one producing the substrate for the function of the other. If the products were induced in close proximity to the genes and could defuse only to the adjacent gene on the *same* chromosome, but not as far as that gene on the *homologous* chromosome, then the *cis* constellation, but not that of the *trans* constellation, would result in complementarity. This interpretation rescued the notion of discrete and distinct genes [Lewis, 1951].

Similar studies, in more "classical" Drosophila genes [Chovnick, 1961; Green, 1963] suggested that they too were divisible into two (or three) distinct subunits, rather than being single genes with mutation-site distributed all over. This notion was finally overcome only with Chovnick *et al.*'s study of the *rosy* gene of Drosophila, in the footsteps Benzer's genetic analysis of the *rII* locus of the T4 bacteriophage of *Escherichia coli* (see below). In these studies *rosy* could be mapped as a continuous linear entity with recombinable mutation sites all along [Chovnick, 1989; Chovnick *et al.*, 1964].

Although the scale of recombination experiments with Drosophila was upgraded considerably, achieving resolution between mutants of 0.01 cM apart and less (cM – centi-Morgan –1% recombination), the introduction of experimental organisms in which more efficient methods for screening for the relevant recombinant progeny was imperative. Pontecorvo utilized auxotrophic mutants (that could not grow on a minimally-defined medium, on which the prototrophic wild-type could grow) of the mold *Aspergillus nidulans.* He hybridized independently obtained mutants of the same gene and screened for the products of rare recombinations by plating the progeny spores on the minimal medium. Inter-allelic (or, inter-pseudo-allelic) recombinants could be discerned from (equally rare) back-mutations by checking for recombination of outside markers — those to the left and right of the experimental target markers. It soon became clear that auxotrophic mutations in given genes

could be mapped into consistent linear sequences [Pontecorvo, 1958]. The genes turned out to be complex, linear structures, more like "molecules" than "atoms" of inheritance.

These results obtained new significance once Watson and Crick presented their model for the genetic material, the double stranded DNA molecule of complimentarily paired nucleotides, adenine with thymine and guanine with cytosine. Muller accepted that "modern evidence supports the conception of a linear arrangement of the genetic material all the way down to its ultimate components, the nucleotides", and "we arrive at a conclusion in harmony with the model that has been presented to us by Watson and Crick" [Muller, 1956]. But he continued to argue for "the case for a segmented chromosome" of discrete and distinct genes, on both experimental — most chromosome breaks are between genes — and theoretical grounds:

> Our best evidence for intragenic mutation ... consists of the evidence in all its great sweep and diversity for the theory of evolution itself, taken in conjunction with the evidence for the existence of individual genes. So long as we retain both concepts we must admit that at the bottom evolution has been built up out of intragenic mutations, together with some superstructure of intergenic structural changes. [Muller, 1956, 134]

As Watson and Crick [1953a] suggested, "many lines of evidence indicate that [DNA] is the carrier of ... the genetic specificity of the chromosomes and thus of the gene itself". Yet, in spite of its elegance, the Watson and Crick model posed some serious problems. Primary among those was that raised by Max Delbrück, of the difficulties in unwinding of long coils of the double helix toward replication and the rewinding of the replicated product (see Holmes [2001]; Kay [2000]). The experimental evidence for the semi-conservative replication of DNA, each double helix of the newly replicated DNA consisting of one helix of the parental double helix, and one newly synthesized [Meselson and Stahl, 1958] — extending the methodology of hybridization to the molecular level – only amplified claims that the chromosomes must be composed of many DNA molecules each unwinding, replicating and rewinding independently, as was implicit in Muller's argument for separate, discrete genes along the chromosomes. At the time Stahl [1961] even suggested that the chromosomes were composed of chains of DNA molecules, each being circular (and containing several genes). However, as noted (see sections 3 and 3.1), the continuity of the DNA molecule of both the bacterial chromosome and that of the eukaryotic chromosomes was eventually established.

The reservations of his mentor Max Delbrück notwithstanding, the new conception of the gene as a stretch of DNA prompted Seymour Benzer to map a gene of the bacteriophage T4, "down to the ground" [Holmes, 2000]. *rII* mutants are *trans* non-complimentary in double infection and do not multiply in the of *E. coli* strain K12(λ) bacteria (whereas they multiply normally in *E. coli* strain S and as *rough* plaques on *E. coli* strain B). Benzer screened mutants belonging to the same complementation unit for "any wild-type recombinants which arise, even in

proportions as low as 10^{-8}" by "crossing" pairs of *rII* mutants in the bacterial strain B, in which they may replicate and recombine, and plating the progeny on a bacterial strain K12(λ) on which mutants cannot produce plaques [Benzer, 1955, 345]. Benzer was not committed to Muller's original conception of the material gene as the composite unit of function, mutation, as well as recombination, and defined the functional unit, the *cistron* (according to the *cis-trans* test, although only many years later the *cis*-combinations of mutants were isolated and directly tested) as not necessarily identical to both the unit of mutation, the *muton*, and the units of recombination, the *recon* [Benzer, 1955; 1957]. Utilizing what appeared upon genetic analysis to be intra-cistronic "deletions" — mutant that did not recombine with (at least) two other mutants that recombined with each other – Benzer greatly improved the efficiency of the mapping analysis and constructed a consistent linear map of the *rII* gene of the bacteriophage T4 that was divisible into two adjacent cistrons. Although it was implicitly presumed that the sited of the mutons and recons would converge to those of individual nucleotides of the cistrons' DNA sequence, neither recons nor mutons corresponded to individual nucleotides. Furthermore, the failure to isolate the protein coded by the *rII* gene did not allow testing the colinearity of the gene's DNA sequence and the protein product's amino-acid sequence. The colinearity of mutations sites in the linear genetic map and changes in amino-acid in the sequences of the protein produced was, however, demonstrated for the tryptophane synthetase B gene, or cistron of *E. coli* [Yanofsky *et al.*, 1964).

The continuity of intra-geneic and inter-genic DNA sequences was indicated by mutations (such as *rII 1589*), which genetically were deletions starting in one cistron and ending at an adjacent one (unless some non-DNA matter was deleted with the DNA sequences). Support for the *structural non*-discreteness of the gene (or cistron) was, however, obtained when it was shown that such a deletion combining what-was-left of the two genes (cistrons) could be "read" *continuously* as long as the welded parts were in the same "reading-frame" (see section 5.1). The gene, or cistron, was shown to be a sequence along the continuous DNA, with no inbuilt structural punctuation signals. To the extent that a unique sequence could be identified, this was by *functional* criteria, such as providing an attachment site for the RNA-polymerase machinery at transcription, or having an open reading frame (ORF) for the ribosomal translation machinery of a nucleotide sequences to an amino-acid sequences.

Arguably, mapping the gene "down to the ground" and demonstrating that the genetic map is collinear with that of the translated polypeptide product of this gene appeared finally to end an epoch: It was not necessary to refer to the gene instrumentally by its product, to infer the genotype from the phenotype, whether eye color or polypeptide. The sequence of nucleotides was the source of "information" for the phenotype. Once the techniques of manipulating DNA were developed, it would be conceivable to refer to a sequence of DNA *per se*, irrespective to its function.

The fundamental unit of classical genetics is an indivisible and ab-

> stract gene. In contrast, the fundamental unit of molecular genetics is a concrete chemical molecule, the nucleotide, with the gene being relegated to a role of a secondary unit aggregate comprising hundreds or thousands of such nucleotides. [Stent, 1970, 910]
>
> ... since it was already taken for granted that the genetic information is held in the form of a particulate nucleotide base sequence in DNA, it now became clear that the *meaning* of the particulate nucleotide base sequence making up a sector of DNA corresponding to a gene could be nothing other than the specification of an amino acid sequence of the corresponding protein molecule. [Stent, 1970, 925]

Attempts to formally reduce the classical gene to the molecular gene [Rosenberg, 1985] (see also Sarkar [1998]; Schaffner [1976]) apparently only supported the claim of discontinuity between the classical gene and the molecular gene. Although another two decades were needed, genetic engineering seemed to have achieved this fit of defining a gene as a sequence of DNA, "which in itself is indeterminate with respect to phenotype" [Moss, 2003, 46]. Ironically, this notion from the DNA-sequence to the trait was called "reverse genetics" though this should have been called "forward genetics", whereas the classical approach of inferring from the phenotype to the genotype is the genuine "reverse genetics".

This dichotomy of the gene as that unit which is identified by a phenotypic "marker" versus the gene as a defined DNA sequence has been developed to its extreme by Lenny Moss:

> The rhetoric of the gene as code and information, ... turns on ... a conflation of two distinctly different meanings of the gene. When scientists and clinicians speak of genes for breast cancer, genes for cystic fibrosis, or genes for blue eyes, they are referring to a sense of a gene defined by its relationship to a phenotype ... and not to a molecular sequence.
>
> It continues to be useful, in some contexts, to employ this usage of the word "gene." To speak of a gene for a phenotype is to speak as if, but *only* as if, it directly determines the phenotype. It is a form of prefomationism but one deployed for the sake of instrumental utility. I call this sense of the gene — Gene P, with the P for preformationist. ...
>
> Quite unlike Gene-P, *Gene-D is defined by its molecular sequence.* A Gene-D is a developmental resource ... which in itself is *indeterminate* with respect to phenotype. To be a Gene-D is to be a transcriptional unit on a chromosome within which are contained molecular template resources. [Moss, 2003, 44-46]

This duality of concepts may have been justified as long as the "classical molecular gene", a unique sequence of DNA that is transcribed and translated into a

unique sequence of amino-acids, was in vague. However, it lost its power with the introduction of the Pandora box of the "bewildering gene" [Falk, 1986]. Once it has been determined that the sequence of DNA that is transcribed to RNA, resembles only vaguely the RNA that is translated to a polypeptide, due to numerous post-translational processes (see section 8), such a definite duality or any other discontinuity in the evolution of the gene concept, seems unwarranted. Clearly there is no one DNA sequence that *on its own*, independently to the context of the product's function, can be unequivocally defined as a gene [Falk, 2000b, 2004].

5.1 The structure of DNA and deciphering of the code

The impact of Muller's research project on genetic analysis can hardly be exaggerated. Muller extended his research program to deduce the physico-chemical nature of genes from the genetic analysis of radiation-induced mutations to include the study of chemically induced mutations. But as pointed out by Charlotte Auerbach [1967] (see also Falk [1986, 152, ftn 72]), one of the pioneers of chemical mutagenesis, he failed in achieving his main object, the elucidation of the physico-chemical nature of the gene by genetic analysis. It took chemical analysis to elucidate the nature of the genetic material.

The history of the nature of the heredity material has been intensively documented (see, e.g., Olby [1974]). Evidence that the hereditary properties of bacteria could be transformed by DNA extracted from other bacteria was of little impression on the genetic research community, not only because bacteria were still considered by many as organisms whose biology profoundly differed from that of higher-organisms, but primarily because the chemistry of DNA was not considered complex enough to provide for the variability that was expected of the hereditary material. Proteins were the only chemicals that showed enough variability and diversity to answer the chemical expectations of genetic analysis. DNAs "boring" repetitive structure could not be the holder of the necessarily rich genetic "information". In retrospect it is amazing to follow the rejections of Avery and others, who persistently indicated that DNA was the carrier of genetic information. Even evidence that the transforming element of bacteria is more than 99% pure deoxyribonucleic acid left the door open for the disbelievers (see Deichmann [2004; Zuckerman and Lederberg [1986]).

Ironically, however, the experiment that turned the balance in convincing the research community of the nature of the genetic material was after all, essentially a biological experiment of bacteriophages (phages) transmitting their DNA rather than their protein to their progeny. When the *Escherichia coli* bacteria were infected with phages whose proteins were S^{35}-labeled, only some of the radioactive sulfur showed up in the progeny phages' proteins, whereas once phages' DNA was labeled with P^{32}, most of the radioactive phosphor of the nucleic acids showed up in their progeny [Hershey and Chase, 1952].

Crick [1958] established on theoretical grounds that the genetic code was one of triplet nucleotides coding for each of the twenty amino-acids (each coding

nucleotide-triplet being a *codon*), and Brenner [1957], upon analyzing dimmers of amino-acids from known peptides, concluded that there were more adjacent amino-acid combinations than and overlapping code will allow. First attempts to decipher the genetic code compared changes of the "wild type" amino-acid at a specific site in the polypeptide of the enzyme tryptophan-synthetase of *E. coli* by two different "mutant" amino-acids, with mutations at different though recombinationally very close sites in the cistron, presumably belonging to the same codon [Hennig and Yanofsky, 1962]. However, the code was deciphered by molecular, rather than by genetic methods. Polyribonucleotide strands of known structure were "fed" to cell extracts and the synthesized polypeptides were extracted and identified [Lengyel *et al.*, 1961; Nirenberg and Matthaei, 1961]. Still, the crucial insight that the genetic code is read in-frame from a given starting point was obtained by genetic analysis of acridine induced mutations in the *rII* cistrons of the T4 bacteriophage [Crick *et al.*, 1961]. It was reasoned that acridine molecules intercalate in the DNA molecule to cause a deletion or an insertion of one nucleotide in the replicating DNA molecule. First order intra-cistronic acriflavin-induced repressor mutants (designated –) of a given *rII* mutant (designated +) were found to be by themselves *rII* mutants. They were followed by second order intra-cistronic repressor mutants of the repressors (designated, again, +) and so on, to higher order suppressors. All suppressor mutations *per se*, i.e., without their respective suppressed mutation, acted as genuine *rII* mutants. Suppressor mutants of different orders were recombined in *cis*. Whereas *cis*-recombinants of a + and — mutation gave as a rule the suppressed, nearly wild-type phenotype, *cis*-recombinants of pairs of +s (or –s) mutants gave the mutant phenotype. However, *cis*-recombining three +s (or –s), even of different orders, gave again a nearly wild-type phenotype. This supported the notion that acroflavine mutants were either deletions or additions of a singe nucleotide (or a small constant multiple of nucleotides) and that the genetic code is a triplet (or a small multiple of a triplet) code, in which reading starts at a given site and continues, comma-less, one triplet (codon) after another, from the initiation codon to the termination codon (three termination codons have been identified among the 64 combinations of four nucleotide triplets, 61 coding for 20 amino acids, one to six codons per amino-acid, the code being redundant). This ingenious analysis as well as the chemical analysis of amino-acid attributions to the codons were elegantly verified for sequences of amino acids of peptide fragments of the gene coding for lysozyme of *E. coli*, in which an acriflavin mutant was induced, and then, its induced *cis*-suppressor was selected. Comparing the amino-acid sequences in the wild-type and between-mutants fragments verified the model of the open reading frame of the code as well as the code assignments of the chemical analysis [Streisinger *et al.*, 1967]. (For a comprehensive review see [Kay, 2000].)

6 FROM GENOTYPES TO PHENOTYPES TO STEREOTYPES

> Tell me your genes
> and I'll tell you
> who you are
> (from a TV program)

A major achievement of Mendel's experimental design was his ability to supersede specific traits and draw analytic, general conclusions on the mechanisms of heredity, irrespective of how the alternative properties unfold. Mendel's method allowed him to ignore the relationship between *Faktorn* and traits. Contrarily, de Vries's *Intracellular Pangenesis* theory was a theory of development and evolution of organisms as assemblies of distinct characteristics. Mendelian *Faktoren* became part of his notion of "unit characters" in his *Mutationstheorie* (see section 2). Thus, research of gene expression from its very beginnings rose beyond the *organism* as the referent and concentrated on the development and function of *unit-characters*, differing from research in embryology that traditionally was an organismic, top-down discipline. However, *Entwicklungsmechanik* was introduced to promote a reductionist notion of development, as heralded in Roux's introductory manifesto to his new *Archiv*:

> In accordance with Spinoza's and Kant's definition of mechanism, every phenomenon underlying causality is designated as a *mechanical phenomenon;* ... Since only phenomena underlying causality are capable of investigation, ... and since the production of form constitutes the essential feature of development, ... *we must trace back each individual formative process to the special combination of energies by which it is conditioned*, or, in other words, to its *modi operandi*; ... Those *modi operandi*, to which we reduce organic formative processes, and hence also the energies which condition them, may be identical with those which underlie inorganic or physico-chemical processes. [Roux, 1904/1986, 108–111]

Still, Roux was very much aware of the complexity of living organisms, and warned that, "Among biologists there is a tendency derived from inorganic sciences, *to regard the hypothetical deductions which appear to us to be the "simplest" as having the greatest probability for the very reason that they seem so simple*" [Roux, 1904/1986, 119]. Thus developmental mechanists were patently in a constant conflict: *Conceptually* they must accept that living organisms are complex entities that cannot be reduced to simple physico-chemical principles, yet *methodologically* they must apply physico-chemical experimental procedures to uncover causes, or *modi operandi*:

> "*Incidit in Scyllam, qui vult vitare Charybdim*" is particularly applicable to the investigator in the field of developmental mechanics. The

> *too simply mechanical* and the *metaphysical conception* represent the Scylla and Charibdis, to steer one's course between which is indeed a difficult task, a task which few have hitherto accomplished. [Roux, 1904/1986, 130]

Facing such an ordeal, it may be no wonder that, to the extent that geneticists were concerned with problems of embryology and development, they preferred the search for phenomenological *modi operandi* of unit traits over that of those of the physico-chemical causes of organismic systems. This shift in perspective is best seen in the "metamorphosis" of Thomas Hunt Morgan, an embryologist who consistently attempted to maintain the center of interest of genetic analysis on problems of development [Allen, 1978]. Morgan overcame his reservations from de Vries' preformationism, once he accepted Johannsen's differentiation between the "transmission conception" of personal qualities and the "genotype conception" of discrete particles that are "'bearers' of special organizing functions in the mechanism of ontogenesis" [Johannsen, 1911, 132]. Yet, he tried desperately to maintain a developmental-system approach of the role of genes and their mutations in development of the Drosophila fly, as evident from the nomenclature of mutants that he introduced: a gene was called by its (presumed) function. But this nomenclature soon became so complicated, and had to be changed with every new mutant, that he reluctantly gave in and adopted a nomenclature that pragmatically expressed instead merely the phenotypic deviation of the mutant from the "wild type", as a designator or marker of that gene [Falk and Schwartz, 1993]. This change of nomenclature inevitably led to a change of emphasis away from the conceptual, functional integrative aspects of the development of traits, and, methodologically, turned attention to the specific phenomenological function of genes.

Johanssen too was quite alarmed by the reductionist implications of the introduction of the notion of the gene into the theory of heredity:

> My term "*gene*" was introduced and generally accepted as a short and unprejudiced word for unit-factors ... but originally I was somewhat possessed with the antiquated morphological spirit in Galton's, Weismann's and Mendel's viewpoints. From a physiological or chemico-biological standpoint ... *there are no unit characters at all!* ...
>
> We may in some way "dissect" the organism descriptively, using all the tricks of terminology as we please. But that is not allowed in Genetical explanation. Here, in the present state of research, we have especially to do with such genotypical units as are separable, ... Certainly by far the most comprehensive and most decisive part of the whole genotype does not seem to be able to segregate in units; and as yet we are mostly operating with "characters," which are rather superficial in comparison with the fundamental Specific or Generic nature of the organism. [Johannsen, 1923, 136–137]

This methodology of viewing the effects of individual genes, relegating interactions to a secondary role, became to an important extent, the conceptual stereotype

of the genotype-phenotype relationship (not only in the eyes of the lay public). Only toward the end of the twentieth century, when developmental biology became, on the one hand more molecular, developmental genetics became also methodologically "physico-chemical" (see Gilbert [1996; 1998; 2000]), while on the other hand, with genetic analysis conceptually becoming more system oriented (see e.g., Oyama *et al.* [2001]), aided by technologies of handling enormous amounts of data of complexity, the two disciplines merge again. It would, however, be a mistake to claim that classical genetic analysis neglected the functional and developmental aspects of organisms: Rather, it concentrated on these aspects from the perspective of specific genes.

Already in 1902, consulting Bateson, Archibald Garrod, who was interested in the metabolic aspects of human conditions, recognized alkaptonuria as an inborn metabolic variant inherited as if it were a recessive Mendelian unit character [Garrod, 1902]. This method of recognizing conditions as "markers" of Mendelian unit characters allowed Garrod within a short period to identify four human unit characters, and point at them as "inborn errors of metabolism" [Garrod, 1908]. (See also Harris [1959].)

One of the early issues of gene function was that of the role of dominance. Whereas Mendel described the dominance of one of the two alternative forms of each of the seven selected factors as a mere empirical fact [Mendel, 1866/1966, 10-11], de Vries considered it as a major principle property: "The lack of transitional forms between any two simple antagonistic characters in the hybrid is perhaps the best proof that such characters are well delimited units" [de Vries, 1900/1966, 110]. De Vries and Hertwig, like Roux and Weismann argued for preformed determinants for all characteristics in the egg, but unlike the latter did not believe that differentiation occurred by the unequal distribution of these determinants to the blastomers. According to de Vries differentiation was achieved by activation-inactivation of latent determinants, or pangenes. Correns could not understand "why de Vries assumes that in all pairs of traits which differentiate two strains, one member must always dominate" [Correns, 1900/1966, 121], and suggested instead a physiological, quantitative, rather than the preformational interpretation for the phenomenon [Correns, 1924a, 330]. But it was Bateson who gave the influential developmental interpretation to dominance in the form of the Presence-and-Absence hypothesis, recessiveness being absence or loss of the *character*. It must, however, be kept in mind that although he did not distinguish between character and factor (or phenotype and genotype), yet, conceptually he considered the development of the organism and he and his coworkers justified the dealing with a single trait by heuristically assigning all the remaining trait to the "background": "[T]he Mendelian contrasting pair, yellow and green might be regarded as presence and absence of yellow *on a basis of* green" [Hurst, 1906, 119 my emphasis]. This exactly was also the policy of Morgan, when he realized that his embryologically-appropriate nomenclature must be neglected for practical reasons: Methododlogically consider the mutant effect of the gene attended, but conceptually mind also the impact of the rest as the "residuum". (For details see

Falk [2001b]; Schwartz [1998].)

As indicated, a crucial conceptual development was Johannsen's distinction between genotype and phenotype [Johannsen, 1909]. At least in theory, it disjoined the discussion of traits from the discussion of the inherited factors, or as put by Johannsen it distinguished between the vernacular "transmission conception" and the "genotypic conception". The unit-character was demoted from its role of a unit of inheritance. This worked two ways: on the one hand this segregation allowed the handling of phenotypic variation more flexibly, rather than the Galtonian dichotomy of "nature *versus* nurture", because the phenotype included both a genetic and a non-genetic component. Even for traits that (phenotypically) vary continuously, the genetic factors may vary discontinuously. This notion had been applied when e.g., East and Nilsson-Ehle extended the Mendelian rules of independent segregation of discrete genes to the continuous variation of "quantitative traits" (see section 2). It was also applied to settle the dispute between Pearson's claim for continuous inheritance based on biometric arguments of phenotypic variation and Bateson's on discontinuous Mendelian genotypic variation (see section 7). On the other hand, by providing non-genotypic "excuses" to the observed phenotypic deviations, it encouraged genetic determinism. Significantly, since the only indicator for the genotype was the phenotypic "marker", the concept of "the gene for" became entrenched as a major determinist notion (see, e.g., Gannett [1999]).

Unfortunately, Johannsen's typologist conception of the genotype, as the basic taxonomic entity [Roll-Hansen, 1978] underpinned his deterministic view of the genotype. When, at the same year that Johannsen published his book, Woltereck tried to emphasize the significance of the environment and its interaction with the genotype in regulating the patterns of individuals' variations [Woltereck, 1909] it was rejected. Woltereck showed that different pure-lines of Daphnia responded differently to the same environmental effects, and that the *Reaktionsnorm* — norm of reaction (NoR) — was specific to the specific line as well as to the specific environmental factor. Conceptually, such specific *Reaktionsnorm* curves of genotypes and environmental conditions would make it impossible to predict the NoR from one set of genotypes and environments to another, or put differently, genotype-phenotype interactions were significant components of the phenotypes. Woltereck's conclusion of genotype-environmental interactions, if accepted, would have countered the central praxis of genetics that uses phenotypic "markers" as indicators, predicting the genotype, which was a basic tool of genetic analysis (as pointed out by Kohler [1994] this was not a major problem with "laboratory-trained" organisms, like Drosophila, which were cultivated under standard "environmental" conditions). Even worse, it suggested the participation of both the genotype and the environment in determining the fitness of the individual in the process of evolution — an explicit Lamarckian notion (Woltereck nursed Lamarckian notions, see Harwood [1996]. As conceived by Hogben already in 1933 this "interdependence" of nature and nurture argued against genetic reductionism [Hogben, 1933]. But genetic analysis was not ripe for such claims at the time, and the genocentric position

was most explicitly reflected by terms such as "penetrance" and "expressivity", the proportion of individuals of a given genotype showing the predicted phenotype and the degree that it showed, respectively, which were introduced by Vogt [1926]. These were considered to be inherent properties of the genes concerned (or of their alleles) [Falk, 2000a; Sarkar, 1999]. A major contributor to this genocentricity was the popularity of the statistical test of the Analysis of Variance that neatly divided the phenotypic variation into a genotypic and an environmental component (with a residual "interaction" component) (see Lewontin [1974] for the analysis of the distortion). As eventually, stated explicitly, the Darwinian notion of natural selection should include the organism shaping its environment as much as that of the environment shapes the organism [Lewontin, 2000].

Contrary to the early attention to internal and external factors that affect sex-determination in various species, and despite of intensive studies such as Goldschmidt's of cytoplasmic factors that vary in the "strength" of sex-determination in the gypsy moth *Lymantria dispar*, Morgan's students, utilized deviations in sex determination of chromosomal aberations in Drosophila to establish a genetic-chromosomally centered theory of sex-determination. According to this the Y-chromosome did not play a role in sexual differentiation (except of conferring fertility on males) — XXY flies were normal and fertile females, whereas XO flies were males but sterile. By producing triploid flies with either 3 or 2 X-chromosomes, Bridges demonstrated that the ratio of X-chromosomes to autosomes (non-sex-chromosomes) determined the flies' sex [Bridges, 1925]. For many years this "balance theory" of sex determination prevailed although exceptions were known, such as in the bisexual plant *Melandrium alba* where the presence of a Y-chromosome was shown to be needed for male-sex determination [Westergaard, 1958]. It was, therefore, somewhat of a shock when in the 1950s it turned out that in humans individuals with XXY chromosomes are phenotypic males (suffering of the Klinefelter syndrome, that includes sterility) and individuals with XO chromosomes are females (suffering from Turner's syndrome).

This notion of an inherent gene action, even when interaction was immanent may best be illustrated by Muller's presentation of "position effect". Whereas Goldschmidt considered position effect to be evidence for the integrity of the chromosome rather than that of discrete genes [Goldschmidt, 1954], Muller considered this very property of being able to interact with neighboring genes, as evidence for the integrity of the discrete genes (see section 5). The advantages of this reductionist research method were considerable. The most direct attempt to abridge the experimental-methodological gap between embryology and genetics was that undertaken by Beadle and Ephrussi [1936], who deliberately applied embryological experimental design to genetic problem-posing. Thus they studied the effects of reciprocal transplantations of imaginal discs (distinct discs of tissue in the body of the larva, each destined to become a specific adult organ) among larvae of Drosophila of various eye-color mutants (and wild type) so as to elucidate the developmental pathway of the production of the eye-color pigments, the ommochromes and the pteridines. In spite of the experimental achievements

with both Drosophila, and the moth *Ephestia kuhnellia* performed by Kühn, it turned out that the systems were too complex for such detailed reductive analysis. Although some embryological and biochemical analysis was further carried out along these lines [Rheinberger, 2000a], Beadle was prompted to look for more basic systems, where it would be possible to establish the effects of single gene. His collaboration with Tatum on the metabolic pathways in the mold *Neurospora crassa* culminated in the reductionist concept of "one gene — one enzyme", that proved to be a productive analysis not only for the elucidation of the function of specific genes, but also *vice versa*, for establishing yet unknown steps in metabolic pathways [Beadle, 1945; Beadle and Tatum, 1941a, 1941b], (see also Schwartz [1998]).

Already in 1929 Sturtevant realized that genetic markers could be extremely helpful tools in developmental analysis. A *claret* eye-color mutant of *Drosophila simulans* proved to induce frequent X-chromosome non-disjunction in early cleavage divisions of the embryo, producing gynandromorphic flies, half of their body being XO, i.e., male, and half being XX, i.e., female [Sturtevant, 1929]. Marking one of the X-chromosomes with a recessive marker like *yellow* produced gynandromorphs in which male parts were marked yellow. Since the demarcation line between the female and male parts varied from one case to another, Sturtevant concluded that the nuclei at the early cleavage divisions were still non-determinate with respect to organ differentiation. Furthermore, using the X-chromosome linked *vermillion* eye color mutant, he found it to be non-autonomous; even when one eye was male and the other female according to the *yellow* marker, both eyes were either vermillion or wild type, according to the genotype of a region of the thorax (which turned out to be the ring-gland). Sturtevant thus made an early genetic analysis fate-mapping of the "focus" of the vermillion eye color of Drosophila.

Genetic analysis of the developmental roles of various specific genes was applied in different organisms. In maize, mutations occurring with abnormally high frequency in genes that affect the development of the seeds' pericarp pigmentation [Emerson, 1914] culminated in McClintock's model of regulatory genes "jumping" to different chromosome sites due to breakage and rejoining cycles [McClintock, 1951]. In the rat the pleiotropic effects on complex phenotype of specific genes was studied by Gruneberg [1938], whereas Glueckson-Waelsch's studies of the embryological effects of single mutations elucidated interactions of basic induction processes in skeleton and urogenital differentiation in the mouse [Gilbert, 1991]. Curt Stern utilized the phenomenon of mitotic crossing-over, inducible by mild doses of X-rays in Drosophila, that yielded in heterozygotes for genetic markers (like *yellow* and *singed*) somatically marked cell "spots" homozygous for the markers, and thus detectable on the background, to follow the site-determination of bristle development in the flies [Stern, 1955]. However, the most systematic embryological research project was that of the genetic control of the developmental course of lethal mutations at specific genes in Drosophila flies [Hadorn, 1961]. Ernst Hadorn and his students studied, among others, the constancy of the developmental fate of imaginal discs of Drosophila: Imaginal discs from larvae were cultured

in vivo by indefinite serial transfers into the abdomen of adult flies. The cultured discs' developmental capacity was examined at will by implanting cultured discs (or parts of the discs) into larvae about to pupate; the disc's developmental fate was examined in the emerging adult's abdomen. As a rule, disc determination was maintained for many generations — a disc determined to become an eye developed into an eye after dozens of serial transfers. However, transdetermination to other developmental fates occurred occasionally, such as an eye disc that developed into an antenna (see e.g., Gehring and Nöthiger [1973]). These studies that demonstrated that the direction of tansdetermination was not random [Kaufman, 1973], and that it was an epigenetic process involving groups of cells that happened to be at a specific site at the right time, independent of their previous determination [Gehring, 1967]. These studies laid the ground for genetic analysis of development, at the molecular level (see, e.g., Ashburner and Wright [1978]).

When Benzer took up Sturtevant's idea of genetic-analysis fate mapping of the Drosophila embryo, locating the foci of behavioral mutants [Hotta and Benzer, 1972] he called the map distances *Sturts* (sixty years after Sturtevant introduced virtual linear linkage maps, the distances of which were measured in *centi-Morgans*). The Sturtevant-Benzer notion was extended further by inducing genetic changes in larval somatic cells — expressed in the adult as mosaic "spots" — at different developmental stages of the larvae. This allowed to follow the dynamic time scale of the hierarchical differential determination of compartments of the imaginal discs [Garcia-Bellido *et al.*, 1979; Garcia-Bellido *et al.*, 1973]. Specific *selector*-genes were identified. These control and regulate differentiation steps and were found to act on groups of cells (polyclones) that happen to be at the right place in the right time, irrespective of their genealogical relationship [Crick and Lawrence, 1975].

One of the serious shortages of genetic analysis of Drosophila was that it concentrated almost exclusively on markers of the adult fly for identifying genes. Of course, there were many lethals who died at embryonic or larval stages of development, but hardly any morphological markers were recognized in Drosophila larvae. This is especially remarkable since so much cytological work was done with the larval polytenic chromosomes. E. B. Lewis had been studying the structure of the *Bithorax* gene complex: *bithorax, Ultrabithorax, postbithorax, Contrabithorax*, etc. (see, e.g., the description of the mutants in Lindsley and Grell [1968] by following the "final", imaginal morphological patterns of the development [Lewis, 1978]. Eventually, by using new microscopic techniques, he noticed numerous larval markers that have been affected by the *Bithorax* genes complex, and extended his genetic analysis to embryonic and larval stages, to come up with a model of genetic control of differentiation, based on genes providing an anterior-posterior gradient in the embryo; the further down the gradient the more genes controlling differentiation of segmentation were activated [Lewis, 1978]. This model, although later modified, heralded a new epoch in genetic analysis of development; it provided a model of differentiation experimentally testable at the molecular level; it extended the power of genetic analysis to the effective study of early develop-

mental stages of the Drosophila as a model organism; and by coinciding with the newly developed molecular "chromosome-walking" method, the bithorax-complex became one of the first to be DNA-sequenced [Bender *et al.*, 1983]. The bithorax complex was actually a complex of homeotic genes, regulatory genes of development, the mutations of which bring about one body part to develop into another (see Section 7). Within a short time it was shown that genetic analysis can be applied successfully to a large scale of genes (known and newly discovered ones) that act at very early stages of embryogeneisis of Drosophila [Nüsslein-Volhard and Wieschaus, 1980] and the era of molecular developmental genetics encompassed an increasing number of animals and plants.

7 POPULATION GENETICS UPHOLDS DARWINISM

Mendel's hypothesis of inheritance of discrete factors that are not diluted should have resolved a major difficulty that Darwin encountered. Shortly after the publication of his *Origin of Species*, in 1867, Fleeming Jenkins showed that, adopting Darwin's theory of inheritance by mixing pangenes, would wash out any achievement of natural selection (see Hull [1973, 302-350]). Hugo de Vries and especially William Bateson, considered Mendel's *Faktoren* as indicated by his hypothesis of inheritance to provide a rational basis for the theory of evolution. Although as early as in 1902 Yule showed that, given small enough steps of variation, the Mendelian model reduces to the biometric claim [Yule, 1902], this was largely ignored in the bitter disputes between the Mendelians and the Biometricians [Provine, 1971], (see Tabery [2004]). Hardy's [1908] proof that in a large population, the proportion of heterozygotes to homozygotes will reach equilibrium after one generation of random mating (provided no mutation or selection interfered), developed in the same year by Weinberg [Stern, 1943], became the basic theorem of population genetics — the Hardy-Weinberg principle. It took, however, another decade for R. A. Fisher to convince that the continuous phenotypic biometric variation reduces to the Mendelian model of polygenes [Fisher, 1918]. Thus, finally the way was cleared to examine the Darwinian theory of natural evolution on the basis of Mendelian genetic analysis, not only *in vivo* but also *in papyro*. As formulated by Fisher in his *fundamental theorem of natural selection*: "The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time" [Fisher, 1930, 37].

Whereas Fisher examined primarily the effects of selection of alleles of single genes in indefinitely large population under the assumption of differences in genotypic fitness, J. B. S. Haldane concentrated on the impact of mutations on the rate and direction of evolution of one or few genes (and the influence of population size) [Haldane, 1990]. Sewall Wright in his models of the dynamics of populations wished to be more "realistic", and stressed the influence of finite population size, the limited gene flow between subpopulations, and the heterogeneity of the habitats in which the population and its subpopulations lived [Wright, 1986].

Experimentally, the main British group, led by E. B. Ford adopted a strict

Mendelian reductionist approach, emphasizing largely the effects of selection on single alleles of specific genes (the evolution of industrial melanism in moths, the evolution of mimicry in African moth species, the evolution of seasonal polymorphisms in snails, etc.). The American geneticists, especially Dobzhansky and his school, concentrated more on problems of whole genotypes, such as speciation (Sturtevant) and chromosomal polymorphisms (Dobzhansky) in Drosophila.

The triumph of reductionist Mendelism was at the 1940s with the emergence of the "New Synthesis" that defined natural populations and the forces that affect their evolution in terms of gene alleles' frequencies [Huxley, 1943]. This notion dominated population genetics for the next decades. Attempts to emphasize the role of non-genetic constraints, such as the anatomical-physiological factors (e.g. by Goldschmidt [1940]), or the environmental (and evolutionary-historical) constraints (for example by Waddington [1957]) were largely overlooked.

The introduction of the analysis of electrophoretic polymorphisms [Hubby and Lewontin, 1966; Lewontin and Hubby, 1966] allowed a molecular analysis of allele variation that was also largely independent of the classical morphological and functional genetic markers (see also Lewontin [1991]). Although genes were still treated as algebraic point entities, inter-genic interacting system, such as "linkage disequilibrium" were considered [Lewontin and Kojima, 1960]. The New Synthesis was, however, seriously challenged when it was realized that a great deal of the variation at the molecular level was determined by stochastic processes, rather than because of differences in fitness [Kimura, 1968; King and Jukes, 1969].

This assault on the notion of the New Synthesis was intensified when, in 1972 Gould and Eldridge, two paleontologists, suggested a model of evolution by "punctuated equilibrium", or long periods of little evolutionary change interspersed with (geologically) relatively short period of fast evolutionary change. Moreover, in the periods of (relatively) fast evolution large one-step "macromutational" changes were established [Eldredge and Gould, 1972]. Although it could be shown that analytically the claims of punctuated equilibrium could be reduced to those of classical population genetics [Charlesworth *et al.*, 1982], these ideas demanded reexamination of the developmental conceptions that, as a rule, could not accept one-step major developmental changes since these called for disturbance in many systems and hence would have caused severe disturbances in developmental and reproductive coordination.

The need to reexamine the reductionist assumptions of genetic population analysis and to pay more consideration to constraints on the genetic determinations of intra- and extra-organismal factors coincided with the resurrection of developmental genetics. However, the major change in the analysis of evolution and development came from the molecular perspective. These allowed first of all detailed upward analysis, from the specific DNA sequences to the early products, rather than the analyses based on end-of-developmental pathway markers. Yet, arguably, the most significant development was the possibility of *in-vitro* DNA hybridization. This molecular extension of genetic analysis *sensu stricto* finally overcame the empirical impossibility to study (most) *in vivo* interspecific hybrids.

The new methods of DNA hybridization had no taxonomic inhibitions whatsoever, and soon hybrid DNA molecules of, say mosquito, human and plant, were common subjects for research. Genetic engineering, which allowed direct genetic comparison between any species and the transfer of genes from one species to individuals of another, unrelated species, prompted the genetic analysis of the evolution of developmental process, or *evo-devo*.

Molecular genetic analysis of homeotic mutants, in which one organ is transformed into the likeness of another, usually a homologous one, revealed stretches of DNA that were nearly identical in other genes with homeotic effects (like the homeobox of some 180 nucleotides, that appear to be involved in when-and-where particular groups of genes are expressed along the embryo axis during development [McGinnis *et al.*, 1984a; McGinnis *et al.*, 1984b]). The method of determining homologies by comparing DNA sequences is nowadays done mainly *in silico*. As suggested many years ago [Ohno, 1970], the abundance of homologous sequences in the same species genome (paralogous sequence that do not necessarily share similar functions any more) or in different species (orthologous sequences that 'usually' have similar functions in different species), indicate that the system's structural and functional organization have been also causal factors rather than merely consequences in the history of the process of evolution.

8 OPENING PANDORA'S BOX: CRACKS IN THE DOGMA

In 1946 Crick formulated the Central Dogma of Genetics:

> I shall argue that the main function of the genetic material is to control (not necessarily directly) the synthesis of proteins... [Crick, 1958, 138]
>
> *The Central Dogma* ... states that once 'information' has passed into protein *it cannot get out again*. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the *precise* determination of sequence, either of bases in the nucleic acid or of amino acid residues in protein. [Crick, 1958, 153]

It is important to quote Crick's full statement, not only because it had obtained such a central role in molecular genetic thinking, but also because often it was quoted in a distorted form. This was the case especially when reverse transcription from DNA to RNA was discovered, when there were assertions such as "Central dogma reversed". Although Crick agreed that the term 'dogma' may have been too strong, obviously "[t]he central dogma was put forward when much of what we now know in molecular genetics was not established. ... In such a situation well constructed theories can play a really useful part in stating problems clearly and thus guiding experiments" [Crick, 1970]. Notwithstanding, the notion that

culminated with the Central Dogma was that genes are repositories of protein-coding sequences. Canonical genes code for proteins.

> No great wonder, then, that many biologists (and journalists) have taken the central dogma to imply that, with very few exceptions, a DNA sequence qualifies as a gene only if it can produce a protein. [Gibbs, 2003, 47]

Crick was careful to state an explicit reservation, "not necessarily directly", when asserting that "the main function of the genetic material is to control the synthesis of proteins". However, the dogmatic role that the Central Dogma fulfilled was to uphold an extreme reductionist conception (see, e.g., Judson [1979, 333–340]), according to which "what holds for *E. coli* holds for the elephant" — a dictum related by some to Crick by others to Monod — could not be maintained any more when molecular geneticists turned their attention to eukaryotes. A persistent problem was the fact that the amount of DNA per cell in eukaryote cells was orders of magnitude greater than that of prokaryotes. Considering that constructing and maintaining cells involves largely household functions, most probably common to eukaryotes and prokaryotes alike, this presented a paradox.

Together with his asserting the Central Dogma, Crick also formulated the theoretical framework based on the experimental evidence, for its materialization. "From many points of view it seems highly likely that the *presence* of RNA is essential for the cytoplasmic protein synthesis" [Crick, 1958, 150]. He predicted this molecule, later known as the messenger-RNA, to be transcribed on the DNA. "Both DNA and RNA have been shown to carry some of the specificity of protein synthesis" [Crick, 1958,151]. He further formulated the adaptor hypothesis,

> It is therefore a natural hypothesis that the amino acid is carried to the template by an 'adaptor' molecule, and that the adaptor is the part which actually fits on the RNA. In its simplest form one would require twenty adaptors, one for each amino acid. [Crick, 1958, 155]

These adaptor molecules, or transfer-RNAs, were shown to be involved in the translation of the nucleotide sequence code into an amino-acid sequence on the ribosomes. The impressive confirmation of this hypothetical model by biochemical experimental evidence (see Rheinberger [1997]), greatly bolstered the genocentric reductionist notion of the Central Dogma.

An early indication of the exhaustion of the notion that DNA sequences qualify as genes only if they can produce a protein was Jacob and Monod's [1961] model of genetic regulation. At the face of it, the model of the genetic regulation of the synthesis of lactose in *Escherichia coli* only bolstered the Central Dogma, upholding the role of RNA as the messenger that conveys information from the DNA sequence to the sequence of amino-acid in protein construction.

Jacob and Monod opened their paper stating:

> According to its most widely accepted modern connotation, the word "gene" designates a DNA molecule whose specific self-replicating structure ... become[s] translated into the specific structure of a polypeptide chain.
>
> This concept of the "structural gene" accounts for the multiplicity, specificity and genetic stability of protein structures, and it implies that such structures are not controlled by environmental conditions or agents. [Jacob and Monod, 1961, 318]

However, the model indicated not only the existence of a new class of genes – or were these really genes? — regulatory genes, but also the existence of polygenic units of transcription: the genes for β-galactosidase, for permease, and for acetylase were all transcribed in a regulated manner into one polycistronic messenger-RNA. To the extent that these genes were meaningful discrete structural entities, discreteness occurred only at the stage of translation of the nucleotide sequences into amino-acid sequences on the ribosomes. In other words, genes were units of the translation complex rather than ones of information encoded in the DNA.

The *in vitro* hybridization of separated DNA strands or of DNA and RNA strands opened new vistas for genetic analysis. Experiments of hybridization of RNA from various tissues of the same organism with its DNA established the notion that tissue differentiation is accompanied by differential activation of genes, and thus — together with the discovery of genetic regulatory systems – allowed a fresh approach to the analysis of molecular control of differentiation. The possibility to hybridize strands of DNA molecules *in vitro* also opened the path to the research of evolution to overcome the barrier of inter-specific hybridization at the molecular level [Hoyer *et al.*, 1964].

The Pandora Box was, however, opened wide in 1968 when Britten and Kohne measured the rate at which the sheared DNA of an organism that was separated into single strands, reassociated to form double strands. The rate of the reassociation or reannealing of the DNA strands is a function of the complexity of the DNA, but much of the DNA of eukaryotes was found to reanneal considerably faster than expected. This had only one explanation, namely, that long sections of the that DNA were redundant and highly repetitive [Britten and Kohne, 1968]. In situ hybridization of radioactively labeled DNA revealed that in the mouse the most highly repetitive sequences (over a million repeats) were concentrated in the heterochromatic segments of the chromosomes, adjacent to the centromeres [Pardue and Gall, 1970]. These are chromosome regions of low genetic informational content, apparently involved in securing regular segregation of chromosomes in the eukaryotic cell division. However, many more repetitive sequences were found to be intercalated in the "genetically active" euchromatic DNA segments. It was suggested that these were parasitic, "selfish DNA" segments [Doolittle and Sapienza, 1980; Orgel and Crick, 1980]. Numerous transposons, or transposable elements — DNA sequences able to insert at many locations in the genome, without sequence relationship with the target locus — were indeed found to be present by

many repeats (the *Alu* family, for example, is represented by $\sim$ 300,000 members, interspersed with nonrepetitive DNA, in the haploid human genome). Although transposons may be of parasitic origins, many turned out to fulfill important (regulatory) functions in the eukaryotic genome. These, and other repeated sequences, comprise a significant proportion of the genome: Only 1.5-2% of the human genome are today considered to code for proteins.

In the years that followed the neat bottom-up picture of the Central Dogma, of DNA transcribed into messenger-RNA that is translated on the ribosomes sequentially and unequivocally into a polypeptide that folds into a protein, had to be greatly modified..

An increasing number of specialized RNA-polymerases were found to be involved in the transcription process. The reductionist negative feedback gene regulation, in which the resources or the products interact automatically with the regulator protein [Monod *et al.*, 1963] was found to be only one component of regulation. Differential activation by the interaction of a very large number of *transcription factors*, many of which attach in rather Baroque constellations to the DNA sequences of the *promoter* regions, provided the framework of cellular driven top-down positive gene regulation by directing the RNA polymerases to the site of transcription initiation. These proteins are crucial to transcription initiation in all eukaryotic cells. More than 2,000 transcription factors are encoded in the human genome. Furthermore, various proteins that affect the binding of the initiation factors, bind to the regulator sequences, upstream of the promoter. The rate of transcription is regulated by interaction with another kind of sequence, sometimes many thousands of nucleotides away from the transcribed sequence, the *enhancer*-sequences that act by binding activator proteins that stimulate (or inhibit) the transcription complex. (For a review, see, e.g., Lewin [2004].)

The concept of the gene as a basic structural DNA entity of heredity got actually the final blow with the discovery that genes of eukaryotes as a rule are not sequences of continuous coding information for translation into polypeptide sequences. Rather, DNA sequences are transcribed into heterogeneous nuclear RNAs, which are processed by a complex, cell-directed splicing mechanism that excerpts several sections (*introns*) from the transcribed sequences, so that only the remaining series of sequences (*exons*) are eventually translated into polypeptide sequences of amino-acids. Furtermore, as a rule, *alternative splicing* allows more than one way to compose a given transcribed sequence to emerge as exons for translation. Often the same primary transcribed sequence may be shuffled into several, sometimes very many alternative RNA sequences. (For a review see Maniatis and Tasic [2002].)

With an increasing number of further editing processes of the pre-tanslational RNAs, such as exon repetition (homotypic trans-splicing), co-transcriptional splicing between two ORFs, anti-sense trans-splicing, alternative trans-splicing from the same or from different chromosomes, antisense overlapping genes, overlapping genes without shared sequences and others, with shared, but alternative reading frames, and many more, the notion of the gene lost its original material concep-

tion completely. It must be conceived as a generic, instrumental unit of function at the molecular level of cellular (or, considering differentiation, even organismal) system's translation processes [Falk, 2000b]. Even the basic claim of the Central Dogma that there cannot be any transfer of amino-acid sequence information from one protein to another must be revised: obviously enzymatic processing of pre-messenger RNA may significantly and specifically modify the message.

Yet, this did not relax the fixation on the reductionist molecular genetic analysis, according to which all "information" for the development of organisms, their growth and function is stored in the DNA and that this is the essence of life. The impressive successes of genetic engineering, which allowed to strikingly manipulate the phenotypes of organisms, only promoted the dogmatic notion of genetic determinism. This was expressed most dramatically by Walter Gilbert, in support of the effort to uncover the full DNA sequence of the genome, which he described as "A vision of the holly Grail" [Gilbert, 1992]. When the Human Genome Project idea started to gain momentum Gilbert presented a CD on which presumably the full sequence of his DNA was stored, asserting: "This is me!"

As has been lately assessed by Mattick the effects of Crick's Dogma will haunt us still for may years to come:

> The central dogma has ... not only been taken to mean that most genes encode proteins, but also that proteins are sufficient in themselves to specify and organize the autopoietic programming of complex biological entities, an assumption that has pervaded molecular biology for decades.
>
> ... a second tier of genetic output and a network of parallel RNA-mediated interactions has evolved in [complex] organisms, which may enable the integration and coordination of sophisticated suites of gene expression required for differentiation and development. [Mattick, 2003, 937]

It was only in recent years, after the intensive efforts of the complete sequencing of the genomes of a large number of species, that the significance of the "non-coding" sequences — sometimes even called "junk sequences" — and especially their role through various kinds of regulatory RNAs (siRNA, miRNA, snoRNA, snRNA, etc.) returned to center-stage (see, e.g., Chen *et al.* [2004]; Dennis [2002]; Enright *et al.* [2003]; Moss and Poethig [2002]). Most of these RNAs act as short segments that hybridize specifically to homologous RNA sequences. Some interfere with translation reactions, others modify or split transfer-RNA or ribonsomal-RNA molecules.

> The failure to recognize the full implications of ... the possibility that the intervening noncoding sequences may be transmitting parallel information in the form of RNA molecules — may well go down as one of the biggest mistakes in the history of molecular biology. (Gibbs [2003] quoting Mattick)

9 CONCLUSION

In an editorial in 1995, "Homage to the chromosome", Joseph Gall wrote:

> Biological organisms, unlike complex inanimate systems, contain information that regulates their day-to-day activities, and, more remarkably, lets them produce new organisms from single cells. That this information resides in the chromosomes was established 75 years ago ... missing from the classical account, however, was any understanding of the nature of the information carried by the chromosomes. That gap was, of course, filled by later spectacular advances in molecular genetics, which showed that a gene is a segment of DNA whose linear sequence of nucleotides specifies the linear sequence of amino acids in the protein. ... But just as a book is more than a random assortment of words, chromosomes are more than simple repositories of gene sequences. They must contain regulatory information for turning genes on and off and they must control their own replication, repair, packaging, as well as the complex movements they carry out during mitosis and meiosis. [Gall, 1995]

Gall here expressed in a nutshell the achievements of genetic analysis together with its methodological successes and epistemological blind alleys.

Genetic analysis in the first decades of the twentieth century had to struggle for its independence [Falk, 1995b]. It achieved this to a large extent by concentrating its efforts on the phenomenological aspects of the mechanical causes of inheritance, by applying a strict reductionist methodology. By and large, genes (and their organization) were treated as intervening variables, or as hypothetical constructs, "because at the level at which the genetic experiment lie, it does not make the slightest difference whether the gene is a hypothetical unit, or whether the gene is a material particle" [Morgan, 1935]. But the wider context was not neglected, whether in the search for their physico-chemical structure and function, or in their role in the evolution of populations and species.

At the time when genetics achieved its secure independence, Morgan asserted:

> The story of genetics has become so interwoven with that of experimental embryology that the two can now to some extent be told as a single story. ... today their interdependence is so obvious that the geneticists takes for granted the main outlines of the facts of embryology, and the embryologist is coming to realize his dependence on the evidence from genetics. [Morgan, 1934, 9]

The course opened by Beadle and Tatum's "one gene — one enzyme" and their analysis of metabolic pathways in *Neurospora* directed wide attention to the study of gene-function in the tradition of genetic analysis of discrete genes, rather than that of dealing with developmental systems. The presentation of the molecular

basis of heredity with Watson and Crick's model of DNA, further directed research studies to gene function rather than to embryonic development and differentiation.

Benzer's detailed mapping of the gene and Yanofsky's demonstration on the colinearity of the gene sequence and the that of protein made the elucidation of the process of gene function more urgent. However, the next step was primarily a task of biochemistry of synthesizing proteins in the test tube [Rheinberger, 1997]. Genetic analysis stepped in again with making sense of the processes of transcription and translation by providing a model of genetic regulation of gene action [Jacob and Monod, 1961].

Many believed at the mid-1960s that molecular genetic analysis has exhausted itself [Stent, 1968], but development and differentiation was not resolved by the molecular genetic analysis of gene regulation and the paradigm of Jacob and Monod [1961]. While molecular biologists encountered the complexity of eukaryotic gene organization, "classical" genetic analysis managed to improve its methodology and increasingly to analyze gene function in a system-context, thus addressing more relevantly the "classical" problems of embryology. It took some time for the molecular biologists to appreciate Lewis's breakthrough in providing a phenomenological model of genetic control of segmental differentiation (see Morange [2000, 196]), but once obtained, genetic analysis of embryological development and differentiation became increasingly molecular.

Evelyn Fox Keller attributes these developments to the change of the mechanistic, lineal mode of thinking to that of the kibernetic, informational feedback era (Keller [1995; 2000; 2002], and many others) that changed the image of the gene as an *acting* agent to that of an *activated* agent. There is no doubt that the "informational" metaphor had an important role in genetic thought. It must however be kept in mind that this metaphor was far from the mathematical-kibernetic notion of information theory, which dealt with the probabilistic reliability of transmission of signs, with no reference to their semantic contents, whereas the information-metaphor of genetic analysis was crucially dependent on the comparison and transmission of *semantic* information, as obtained by the methodology of hybridization, whether that of living organisms or that of DNA and/or RNA nucleotide sequences. More importantly, Keller ignores the internal developments in the conceptions and techniques of the sciences involved. Both Development Systems Analysis (see Oyama [1985]), and the notion of Punctuated Evolution (see Eldredge and Gould [1972]), which played major roles in the new organismal-look, were engendered far away from the foci of molecular-biology, by avid opponents of the information notion. By the time computing algorithms and machines for analysis of very large number of data became available the information metaphor was long forgotten in molecular biology.

Keller is right in pointing at Goldschmidt, who "was typically grandiose, leaning always toward overarching generalization", unrelentingly insisted on embryological development as systems of interacting and coordinated reactions: "[H]is search was precisely for the dynamic properties of such systems (*Zusammenspiel der Reaktionen*). To Goldschmidt gene action meant that genes were both catalysts

and catalyzed, actors and 'reacting substances'" [Keller, 1995, 15-16]. However, to the "American geneticist" Sturtevant's rephrasing of the problems in terms of "chain of reactions into their individual links", had an immense appeal, because, as Keller says, "Once the problem of development is translated into the question of how genes produce their effects, the task is immediately — and almost miraculously — simplified" [Keller, 1995, 16]. The problem was not that of switching the notion from "gene action" to that of "gene activation". It was one of overcoming the reductionist paradigm of genetic analysis that was immanent to the genetic theory of inheritance for de Vries, but largely pragmatic and instrumental to Morgan and his associates.

If genetic determinism became stereotypic, this was to a large extent due to its experimental success. But genetic analysis advanced toward a system's perspective from within. To a large extent, the sheer amount of reductionist data that needed integration led to the sublimation of instrumental reductionism of genetic analysis. Methods were developed toward the beginning of the twenty-first century in relation to the sequencing effort of the complete human genome (see, e.g., Sulston and Ferry [2002]). Many were based on the principles of analysis of the hybridability of nucleotide sequences. The price for such collections of huge amounts of data was inevitably the reduced reliability of individual data. This was overcome in the genome sequencing effort by repeatedly sequencing of any segment. However, for other purposes, like that of building "proteomes" (the total yield of proteins synthesized by cells) or the "interactome" (the total pattern of protein-interactions in given cells) (see, e.g., Perkel [2004]) the lack of accuracy of individual data was often amply compensated for by the algorithms developed for looking for bulk-effects (based on tens of thousands of individual data) of patterns of production and interaction.

BIBLIOGRAPHY

[Allen, 1966] G. E. Allen. Thomas Hunt Morgan and the problem of sex determination, 1903-1910. *Proceedings of the American Philosophical Society*, 110: 48–57, 1966.

[Allen, 1978] G. E. Allen. *Thomas Hunt Morgan: The Man and His Science*. Princeton, NJ: Princeton University Press, 1978.

[Altenburg and Muller, 1920] E. Altenburg and H. J. Muller. The genetic basis of truncate wing, – an inconstant and modifiable character in Drosophila. *Genetics*, 5: 1–59, 1920.

[Amundson, 2005] R. Amundson. *The Changing Role of the Embryo in Evolutionary Thought: Roots of Evo-Devo*. Cambridge: Cambridge University Press, 2005.

[Anderson, 1925] E. G. Anderson. Crossing over in a case of attached X chromosomes in *Drosophila melanogaster*. *Genetics*, 10: 403–417, 1925.

[Ashburner and Wright, 1978] M. Ashburner and T. R. F. Wright, (eds.). *The Genetics and Biology of Drosophila* (Vol. 2c). London, New York, San Francisco: Academic Press, 1978.

[Auerbach, 1967] C. Auerbach. The chemical production of mutations. *Science*, 158: 1141–1147, 1967.

[Bateson, 1894] W. Bateson. *Materials for the Study of Variation*. London: Macmillan, 1894.

[Bateson, 1928] W. Bateson. A suggestion as to the nature of the "walnut" comb in fowls. In R. C. Punnett (ed.), *Scientific Papers of William Bateson*, Vol. II, pages 135–138. Cambridge: The University Press, 1928. First published 1905.

[Bateson and Punnett, 1911] W. Bateson and R. C. Punnett. On gametic series involving reduplication of certain terms. *Journal of Genetics*, 1(4): 293–302, 1911.

[Beadle, 1945] G. W. Beadle. Biochemical genetics. *Chemical Review*, 37: 15–96, 1945.
[Beadle and Ephrussi, 1936] G. W. Beadle and B. Ephrussi. The differentiation of eye pigments in Drosophila as studied by transplantation. *Genetics*, 21(3): 225–247, 1936.
[Beadle and Tatum, 1941a] G. W. Beadle and E. L. Tatum. Experimental control of development and differentiation: Genetic control of developmental reactions. *The American Naturalist*, 75: 107–116, 1941a.
[Beadle and Tatum, 1941b] G. W. Beadle and E. L. Tatum. Genetic control of biochemical reaction in *Neurospora. Proceedings of the National Academy of Science, Washington*, 27: 499–506, 1941b.
[Belling, 1928] J. Belling. A working hypothesis for segmental interchange between homologous chromosomes in flowering plants. *University of California Publications in Botany*, 14(8): 283–291.
[Bender *et al.*, 1983] W. Bender, M. Akam, F. Karch, P. Beachy, M. Pfeifer, P. Spierer, E. B. Lewis, and D. Hogness. Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science*, 221: 23–29, 1983.
[Benzer, 1955] S. Benzer. Fine structure of a genetic region in bacteriophage. *Proceedings of the National Academy of Science, Washington*, 41(6): 344–354, 1955.
[Benzer, 1957] S. Benzer. The elementary units of heredity. In B. Glass and W. D. McElroy (eds.), *The Chemical Basis of Heredity*, pages 70-93. Baltimore, MD: Johns Hopkins Press, 1957.
[Beurton *et al.*, 2000] P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.). *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge and New York: Cambridge University Press, 2000.
[Blackburn, 1991] E. H. Blackburn. Structure and function of telomeres. *Nature*, 350: 569–573, 1991.
[Blakeslee, 1922] A. F. Blakeslee. Variations in *Datura* due to changes in chromosome number. *The American Naturalist*, 56: 16–31, 1922.
[Boveri, 1902] T. Boveri. Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns. *Verhandlungen der physikalische-medizinische Gesellschaft, Würzburg, N.F.*, 35: 67–90, 1902.
[Brannigan, 1979] A. Brannigan. The reification of Mendel. *Social Studies of Science*, 9: 423–454, 1979.
[Brenner, 1957] S. Brenner. On the impossibility of all overlapping triplet codes in information transfer from nucleic acid to proteins. *Proceedings of the National Academy of Science, Washington*, 43(8): 687–694, 1957.
[Brenner *et al.*, 1990] S. Brenner, W. Dove, I. Herskowitz, and R. Thomas. Genes and development: Molecular and logical themes. *Genetics*, 126(3): 479–486, 1990.
[Bridges, 1916] C. B. Bridges. Non-disjunction as proof of the chromosome theory of heredity. *Genetics*, 1: 1–52, 107–163, 1916.
[Bridges, 1917] C. B. Bridges. Deficiency. *Genetics*, 2: 445–465, 1917.
[Bridges, 1925] C. B. Bridges. Sex in relation to chromosomes and genes. *The American Naturalist*, 59: 127–137, 1925.
[Bridges, 1935] C. B. Bridges. Salivary chromosome maps. *The Journal of Heredity*, 26: 60–64, 1935.
[Bridges, 1938] C. B. Bridges. A revised map of the salivary gland X-chromosome. *The Journal of Heredity*, 29: 11–13, 1938.
[Bridges and Anderson, 1925] C. B. Bridges and E. G. Anderson. Crossing over in the X chromosome of triploid females of *Drosophila melanogaster*. *Genetics*, 10: 418–441, 1925.
[Bridges and Bridges, 1939] C. B. Bridges and P. N. Bridges. A new map of the second chromosome. *The Journal of Heredity*, 30: 475–477, 1939.
[Britten and Kohne, 1968] R. J. Britten and D. E. Kohne. Repeated sequences in DNA. *Science*, 161(3841): 529–540, 1968.
[Burian, 2004] R. M. Burian. Molecular epigenesis, molecular pleiotropy, and molecular gene definitions. *History and Philosophy of the Life Sciences*, 26(1): 59–80, 2004.
[Cairns, 1963] J. Cairns. The bacterial chromosome and its manner of replication as seen in autoradiography. *Journal of Molecular Biology*, 6(3): 208–213, 1963.
[Carlson, 1959] E. A. Carlson. Allelism, pseudoallelism, and complementation at the *dumpy* locus in *D. melanogaster*. *Genetics*, 44: 347–373, 1959.
[Carothers, 1913] E. E. Carothers. The Mendelian ratio in relation to certain Orthopteran chromosomes. *Journal of Morphology*, 24: 487–511, 1913.

[Castle, 1906] W. E. Castle. Yellow mice and gametic purity. *Science n.s.*, 24: 275–281, 1906.

[Castle, 1919a] W. E. Castle. Piebald rats and the theory of genes. *Proceedings of the National Academy of Science, Washington*, 5: 126–130, 1919.

[Castle, 1919b] W. E. Castle. Is the arrangement of the genes in the chromosome linear? *Proceedings of the National Academy of Science, Washington*, 5(2): 25–32, 1919.

[Cech, 2004] T. R. Cech. Beginning to understand the end of the chromosome. *Cell*, 116: 273–279, 2004.

[Charlesworth *et al.*, 1982] B. Charlesworth, R. Lande, and M. Slatkin. A neo-Darwinian commentary on macroevolution. *Evolution*, 36: 474–498, 1982.

[Chen *et al.*, 2004] C.-Z. Chen, L. Li, H. F. Lodish, and D. P. Bartel. MicroRNAs modulate hematopoietic lineage differentiation. *Science*, 303: 83–86, 2004.

[Chovnick, 1961] A. Chovnick. The garnet locus in *Drosophila melanogaster*. I. Pseudoallelism. *Genetics*, 46: 493–507, 1961.

[Chovnick, 1989] A. Chovnick. Intragenic recombination in Drosophila: The *rosy* locus. *Genetics*, 123(4): 621–624, 1989.

[Chovnick *et al.*, 1964] A. Chovnick, A. Schalet, R. P. Kernaghan, and M. Kraus. The rosy cistron in *Drosophila melanogaster* genetic fine structure analysis. *Genetics*, 50: 1254–1259, 1964.

[Cock, 1973] A. G. Cock. William Bateson, Mendelism and biometry. *Journal of the History of Biology*, 6(1): 1–36, 1973.

[Coleman, 1970] W. Coleman. Bateson and chromosomes: Conservative thought in science. *Centaurus*, 15(3-4): 228–314, 1970.

[Cooper, 1948] K. W. Cooper. A new theory of non-disjunction in female *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, Washington*, 34: 179–187, 1948.

[Correns, 1900/1966] C. Correns. G. Mendel's Regel über das Verhalten der Nachkommenschaft der Rassenbastarde. *Berichte der deutschen botanischen Gesellschaft*, 18: 158–168, 1900. Translated and reprinted as 'G. Mendel's law concerning the behavior of progeny of varietal hybrids' in C. Stern and E. Sherwood (eds.), *The Origins of Genetics: A Mendel Sourcebook*, San Francisco, W. H. Freeman, 1966.

[Correns, 1902] C. Correns. Scheinbare Ausnahmen von der Mendel'schen Spaltungregel für Bastarde. *Berichte der deutschen botanischen Gesellschaft*, 20(3): 159–172, 1902.

[Correns, 1924a] C. Correns. Über die dominierenden Merkmale der Bastarde. In *Carl Correns: Gesammelte Abhandlungen zur Vererbungswissenschaft aus periodischen Schrifren, 1899-1924*, pages 329–342, Berlin: Julius Springer, 1924. First published 1903.

[Correns, 1924b] C. Correns. *Carl Correns: Gesammelte Abhandlungen zur Vererbungswissenschaft aus periodischen Schrifren, 1899-1924*, Berlin: Julius Springer, 1924.

[Creighton and McClintock, 1931] H. B. Creighton and B. McClintock. A correlation of cytological and genetical crossing-over in *Zea mays*. *Proceedings of the National Academy of Science, Washington*, 17: 485–497, 1931.

[Crick, 1958] F. H. C. Crick. On protein synthesis, *Symposium of the Society for Experimental Biology. The Biological Replication of Macromolecules*, Vol. 12, pages 138–163. Cambridge: Cambridge University Press, 1958.

[Crick, 1970] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258): 561–563, 1970.

[Crick *et al.*, 1961] F. H. C. Crick, L. Barnett, S. Brenner, and R. J. Watts-Tobin. General nature of the genetic code for proteins. *Nature*, 192(4809): 1227–1232, 1961.

[Crick and Lawrence, 1975] F. H. C. Crick and P. A. Lawrence. Compartments and polyclones in insect development. *Science*, 189: 340–347, 1975.

[Darbishire, 1902] A. D. Darbishire. Note on the results of crossing Japanese Waltzing mice with Eruopean Albino races. *Biometrika*, 2: 101–104, 1902.

[Darden, 1977] L. Darden. William Bateson and the promise of Mendelism. *Journal of the History of Biology*, 10: 87–106, 1977.

[Dawkins, 1976] R. Dawkins. *The Selfish Gene*. Oxford: Oxford University Press, 1976.

[de Vries, 1889] H. de Vries. *Intracellulare Pangenesis*. Jena: Gustav Fischer, 1889.

[de Vries, 1900/1966] H. de Vries. Das Spaltungsgesetz der Bastarde. *Berrichte der deutschen botanischen Gesellschaft*, 18: 83–90, 1900. Translated and reprinted as 'The law of segregation in hybrids', in C. Stern and E. Sherwood (eds.), *The Origins of Genetics: A Mendel Sourcebook*, San Francisco, W. H. Freeman, 1966.

[de Vries, 1902–3] H. de Vries. *Die Mutationstheorie: Versuche und Beobachtungen über die Entstehung von Arten im Pflanzenreich*. Leipzig: Von Veit, 1902-3.

[de Vries, 1912] H. de Vries. *Species and Varieties: Their Origin by Mutation* (3rd, corrected and revised ed.). Chicago: The Open Court Publishing Company, 1912. First published 1904.

[Deichmann, 2004] U. Deichmann. Early responses to Avery *et al.*'s paper on DNA as hereditary material. *Historical Studies in the Physical Sciences*, 34(2): 207–232, 2004.

[Demerec and Hoover, 1936] M. Demerec and M. E. Hoover. Three related X-chromosome deficiencies in Drosophila. *The Journal of Heredity*, 27: 206–212, 1936.

[Dennis, 2002] C. Dennis. Gene regulation: The brave new world of RNA. *Nature*, 418(6894): 122–124, 2002.

[Dietrich, 2000] M. R. Dietrich. From gene to genetic hierarchy: Richard Goldschmidt and the problem of the gene. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge and New York: Cambridge University Press, pages 91–114, 2000.

[Dobzhansky, 1929] T. Dobzhansky. Genetical and cytological proof of translocations involving the third and the fourth chromosomes of *Drosophila melanogaster*. *Biologisches Zentralblat*, 49: 408–419, 1929.

[Dobzhansky, 1931] T. Dobzhansky. Translocations involving the second and the fourth chromosomes of *Drosophila melanogaster*. *Genetics*, 16: 629–658, 1931.

[Doolittle and Sapienza, 1980] F. W. Doolittle and C. Sapienza. Selfish genes, the phenotype paradigm and genome evolution. *Nature*, 284: 601–603, 1980.

[Dunn, 1965] L. C. Dunn. *A Short History of Genetics*. New York: McGraw Hill, 1965.

[East, 1912] E. M. East. The Mendelian notation as a description of physiological facts. *The American Naturalist*, 46: 633–655, 1912.

[Eldredge and Gould, 1972] N. Eldredge and S. J. Gould. Punctuated equilibrium: An alternative to phyletic gradualism. In T. J. M. Schopf (ed.), *Models of Paleobiology*. San Francisco: Freeman, Cooper & Co, pages 82–115, 1972.

[Emerson, 1914] R. A. Emerson. The inheritance of a recurring somatic variation in variagated ears of maize. *The American Naturalist*, 48: 87–115, 1914.

[Enright *et al.*, 2003] A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, and D. S. Marks. MicroRNA targets in *Drosophila*. *Genome Biology*, 5(1): R1.1-14, 2003.

[Falk, 1955] R. Falk. Some preliminary observations on crossing-over and non-disjunction. *Hereditas*, 41: 376–383, 1955.

[Falk, 1986] R. Falk. What is a gene? *Studies in the History and Philosophy of Science*, 17(2): 133–173, 1986.

[Falk, 1991] R. Falk. The dominance of traits in genetic analysis. *Journal of the History of Biology*, 24(3): 457–484, 1991.

[Falk, 1995a] R. Falk. The manifest and the scientific. In S. Maasen, E. Mendelsohn, and P. Weingart (eds.), *Biology as Society, Society as Biology: Metaphors*, pages 57–79). Dordrecht, Holland: Kluwer Academic Publishers, 1995a.

[Falk, 1995b] R. Falk. The struggle of genetics for independence. *Journal of the History of Biology*, 28(2): 219–246, 1995b.

[Falk, 1997] R. Falk. Muller on development. *Theory in Biosciences*, 116(4): 349–366, 1997.

[Falk, 2000a] R. Falk. Can the norm of reaction save the gene concept? In R. S. Singh, C. B. Krimbas, D. B. Paul, and J. Beatty (eds.), *Thinking About Evolution: Historical, Philosophical and Political Perspectives*, Vol. 2, pages 119–140. Cambridge: Cambridge University Press, 2000a.

[Falk, 2000b] R. Falk. The gene –a concept in tension. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, pages 317–348. Cambridge and New York: Cambridge University Press, 2000b.

[Falk, 2001a] R. Falk. Mendel's hypothesis. In G. E. Allen and R. M. MacLeod (eds.), *Science, History and Social Activism: A Tribute to Everett Mendelsohn*, pages 77–86. Dordrecht: Kluwer Academic Publishers, 2001a.

[Falk, 2001b] R. Falk. The rise and fall of dominance. *Biology and Philosophy*, 16(3): 285–323, 2001b.

[Falk, 2003] R. Falk. Linkage: From particulate to interactive genetics. *Journal of the History of Biology*, 36(1): 87–117, 2003.

[Falk, 2004] R. Falk. Long Live the Genome! So should the gene. *History and Philosophy of the Life Sciences*, 26(1): 105–121, 2004.

[Falk and Sarkar, 1991] R. Falk and S. Sarkar. The real objective of Mendel's paper. *Biology and Philosophy*, 6: 447–451, 1991.

[Falk and Schwartz, 1993] R. Falk and S. Schwartz. Morgan's hypothesis of the genetic control of development. *Genetics*, 134: 671–674, 1993.

[Fincham *et al.*, 1991] J. R. S. Fincham, P. R. Day, and A. Radford. *Fungal Genetics* (3rd ed.). London: Blackwell, 1991. First edition 1979.

[Fisher, 1918] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society, Edinburgh*, 52: 399–433, 1918.

[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Clarendon, 1930.

[Fisher, 1936] R. A. Fisher. Has Mendel's work been rediscovered? *Annals of Science*, 1: 115–137, 1936.

[Gall, 1995] J. G. Gall. Homage to the chromosome (editorial). *Science*, 270(5242), 1551, 1995.

[Gannett, 1999] L. Gannett. What's the cause? The pragmatic dimension of genetic explanation. *Biology and Philosophy*, 14(3): 349–374, 1999.

[Garcia-Bellido *et al.*, 1979] A. Garcia-Bellido, P. A. Lawrence, and G. Morata. Compartments in animal development. *Scientific American*, 241: 102–110, 1979.

[Garcia-Bellido *et al.*, 1973] A. Garcia-Bellido, P. Rippoll, and G. Morata. Developmental compartmentaliztion of the wing disc of *Drosophila*. *Nature, New Biology*, 245: 251–253, 1973.

[Garrod, 1902] A. E. Garrod. The incidence of alkaptunuria: A study in chemical individuality. *Lancet*, 2: 1616–1620, 1902.

[Garrod, 1908] A. E. Garrod. *Inborn Errors of Metabolism*, (2nd edition 1923 ed.). Oxford: Oxford University Press, 1908.

[Gayon, 2000] J. Gayon. From measurement to organization: A philosophical scheme for the history of the concept of heredity. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, pages 69–90. Cambridge and New York: Cambridge University Press, 2000.

[Gehring, 1967] W. J. Gehring. Clonal analysis of determination dynamics in cultures of imaginal discs of *Drosophila melanogaster*. *Developmental Biology*, 16: 438–456, 1967.

[Gehring and Nöthiger, 1973] W. J. Gehring and R. Nöthiger. The imaginal discs of *Drosophila*. In S. J. Counce, C. H. Waddington (eds.), *Developmental Systems*, Vol. 2, pages 211–290. London: Academic Press, 1973.

[Gibbs, 2003] W. W. Gibbs. The unseen genome: Gems among the junk. *Scientific American*, 289(5): 46–53, 2003.

[Gilbert, 1978] S. F. Gilbert. The embryological origins of the gene theory. *Journal of the History of Biology*, 11(2): 307–351, 1978.

[Gilbert, 1991] S. F. Gilbert. Induction and the origins of developmental genetics. In S. Gilbert (ed.), *Developmental Biology. A Comprehensive Synthesis. A Conceptual History of Modern Embryology*, pages 181–206. Baltimore and London: Johns Hopkins University Press, 1991.

[Gilbert, 1996] S. F. Gilbert. Enzymatic adaptation and entrance of molecular biology into embryology. In S. Sarkar (ed.), *The Philosophy and History of Molecular Biology: New Perspectives*, pages 101–123. Dordrecht: Kluwer Academic Publishers, 1996.

[Gilbert, 1998] S. F. Gilbert. Bearing crosses: A historiography of genetics and embryology. *American Journal of Medical Genetics*, 76(2): 168–182, 1998.

[Gilbert, 2000] S. F. Gilbert. Genes classical and genes developmental: The different use of genes in evolutionary syntheses. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, pages 178–192. Cambridge and New York: Cambridge University Press, 2000.

[Gilbert *et al.*, 1996] S. F. Gilbert, J. M. Opitz, and R. A. Raff. Resynthesizing evolutionary and developmental biology. *Developmental Biology*, 173: 357–372, 1996.

[Gilbert, 1992] W. Gilbert. A vision of the grail. In D. J. Kevles and L. Hood (eds.), *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, pages 83–97. Cambridge, MA: Harvard University Press, 1992.

[Goldschmidt, 1917] R. Goldschmidt. Crossing over ohne Chiasmatypie? *Genetics*, 2: 82–95, 1917.

[Goldschmidt, 1938] R. Goldschmidt. The theory of the gene. *Scientific Monthly*, 45: 268–273, 1938.

[Goldschmidt, 1940] R. Goldschmidt. *The Material Basis of Evolution*. New Haven: Yale University Press, 1940.

[Goldschmidt, 1954] R. Goldschmidt. Different philosophies of genetics. *Science*, 119: 703–710, 1954.

[Goldschmidt, 1955] R. Goldschmidt. *Theoretical Genetics*. Berkeley and Los Angeles: University of California Press, 1955.

[Green, 1963] M. M. Green. Interallelic complementation and recombination at the *rudimentary* wing locus in *Drosophila melanogaster*. *Genetica*, 34: 242–253, 1963.

[Griesemer, 2000] J. R. Griesemer. Reproduction and reduction of genetics. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, pages 240–285. Cambridge and New York: Cambridge University Press, 2000.

[Gruneberg, 1938] H. Gruneberg. An analysis of the "pleiotropic" effects of a new lethal mutation in the rat (*Mus norvegicus*). *Proceedings of the Royal Society, Lundon*, series B, 125: 123–144, 1938.

[Hadorn, 1961] E. Hadorn. *Developmental Genetics and Lethal Factors*. New York: John Wiley, 1961.

[Haldane, 1990] J. B. S. Haldane. *The Causes of Evolution*. Princeton, NJ: Princeton University Press, 1990. First published 1932.

[Hardy, 1908] G. H. Hardy. Mendelian proportions in a mixed population. *Science*, 28: 49–50, 1908.

[Harris, 1959] H. Harris. *Human Biochemical Genetics*. Cambridge: Cambridge University Press, 1959.

[Harwood, 1996] J. Harwood. Weimar culture and biological theory: A study of Richard Woltereck (1877–1944). *History of Science*, 34: 347–377, 1996.

[Hayes, 1964] W. Hayes. *The Genetics of Bacteria and Their Viruses: Studies in Basic Genetics and Molecular Biology*. New York: John Wiley, 1964.

[Heitz and Bauer, 1933] E. Heitz and H. Bauer. Beweise für die Chromosomennatur der Kernschleifen in den Knäuelkernen von *Bibio hortulanus*. *Zeitschrift für Zellforschung und mikroskopische Anatomie*, 17: 76–82, 1933.

[Hennig and Yanofsky, 1962] U. Hennig and C. Yanofsky. An alteration in the primary structure of a protein predicted on the basis of genetic recombination data. *Proceedings of the National Academy of Science, Washington*, 48: 183–190, 1962.

[Hershey and Chase, 1952] A. D. Hershey and M. Chase. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology*, 36(1): 39–56, 1952.

[Hogben, 1933] L. Hogben. *Nature and Nurture*, New York: W. W. Norton, 1933.

[Holliday, 1964] R. Holliday. A mechanism for gene conversion in fungi. *Genetical Research*, 3: 472–486, 1964.

[Holmes, 2000] F. L. Holmes. Seymour Benzer and the definition of the gene. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution*, pages 115–155. Cambridge: Cambridge University Press, 2000.

[Holmes, 2001] F. L. Holmes. *Meselson, Stahl, and the Replication of DNA: A History of "The Most Beautiful Experiment in Biology"*. New Haven, CT: Yale University Press, 2001.

[Hotta and Benzer, 1972] Y. Hotta and S. Benzer. Mapping of behavior in *Drosophila melanogaster*. *Nature*, 240: 527–535, 1972.

[Hoyer *et al.*, 1964] B. H. Hoyer, B. J. McCarthy, and E. T. Bolton. A molecular approach in the systematics of higher organisms. *Science*, 144: 959–967, 1964.

[Hubby and Lewontin, 1966] J. L. Hubby and R. C. Lewontin. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics*, 54: 577–594, 1966.

[Huberman and Riggs, 1968] J. A. Huberman and D. A. Riggs. On the mechanism of DNA replication in mammalian chromosomes. *Journal of Molecular Biology*, 32: 327–341, 1968.

[Hull, 1973] D. L. Hull. *Darwin and His Critics*. Chicago and London: The University of Chicago Press, 1973.

[Hurst, 1906] C. C. Hurst. Mendelian characters in plants and animals. *London: Royal Horticultural Society*, pages 114–129, 1906.

[Huxley, 1943] J. Huxley. *Evolution: The Modern Synthesis*. New York: Harper & Row, 1943.

[Jacob and Monod, 1961] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3: 318–356, 1961.
[Jacob and Wollman, 1961] F. Jacob and E. L. Wollman. *Sexuality and the Genetics of Bacteria*. New York: Academic Press, 1961.
[Janssens, 1909] F. A. Janssens. La Théorie de la chiasmatypie. Nouvelle interprétation des cinèses de maturation. *La Cellule*, 25: 389–411, 1909.
[Johannsen, 1909] W. Johannsen. *Elemente der exakten Erblichkeitslehre*. Jena: Gustav Fischer.
[Johannsen, 1911] W. Johannsen. The genotype conception of heredity. *The American Naturalist*, 45(531): 129–159, 1911.
[Johannsen, 1923] W. Johannsen. Some remarks about units in heredity. *Hereditas*, 4: 133–141, 1923.
[Judson, 1979] H. F. Judson. *The Eighth Day of Creation: Makers of Revolution in Biology*. New York: Simon and Schuster, 1979.
[Kaufman, 1973] S. A. Kaufman. Control circuits for determination and transdetermination. *Science*, 181: 310–318, 1973.
[Kavenoff and Zimm, 1973] R. Kavenoff and B. H. Zimm. Chromosome-sized DNA molecules from *Drosophila*. *Chromosoma*, 41: 1–27, 1973.
[Kay, 2000] L. E. Kay. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford, CA: Stanford University Press, 2000.
[Keller, 1995] E. F. Keller. *Refiguring Life: Metaphors of Twentieth-Century Biology*. New York: Columbia University Press, 1995.
[Keller, 2000] E. F. Keller. *The Century of the Gene*. Cambridge, MA: Harvard University Press, 2000.
[Keller, 2002] E. F. Keller. *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines*. Cambridge, MA: Harvard University Press, 2002.
[Kimura, 1968] M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217: 624–626, 1968.
[King and Jukes, 1969] J. L. King and T. H. Jukes. Non-Darwinian evolution. *Science*, 164: 788–798, 1969.
[Kohler, 1994] R. E. Kohler. *Lords of the Fly. Drosophila Genetics and Experimental Practice*. Chicago: University of Chicago Press, 1994.
[Kornberg, 1974] R. D. Kornberg. Chromatin structure: A repeating unit of histones and DNA. *Science*, 184: 868–871, 1974.
[Kostoff, 1930] D. Kostoff. Discoid structure of the spireme. *The Journal of Heredity*, 21(7): 323–324, 1930.
[Kottler, 1979] M. J. Kottler. Hugo de Vries and the rediscovery of Mendel's laws. *Annals of Science*, 36: 517–538, 1979.
[Lederberg, 1947] J. Lederberg. Gene recombination and linked segregation in *Escherichia coli*. *Genetics*, 32: 505–525, 1947.
[Lederberg, 1991] J. Lederberg. The gene (H. J. Muller 1947). *Genetics*, 129: 313–316, 1991.
[Lederberg *et al.*, 1951] J. Lederberg, E. M. Lederberg, N. D. Zinder, and E. R. Lively. Recombination analysis of bacterial heredity. *Cold Spring Harbor Symposia on Quantitative Biology*, 16: 413–443, 1951.
[Lederberg and Tatum, 1946] J. Lederberg and E. L. Tatum. Gene recombination in *Escherichia coli*. *Nature*, 158: 558, 1946.
[Lengyel *et al.*, 1961] P. Lengyel, J. F. Speyer, and S. Ochoa. Synthetic polynucleotides and the amino acid code. *Proceedings of the National Academy of Sciences, Washington*, 47: 1936–1942, 1961.
[Lenoir, 1982] T. Lenoir. *The Strategy of Life*. Chicago: University of Chicago Press, 1982.
[Lewin, 2004] B. Lewin. *Genes VIII*. Upper Saddle River, NJ: Pearson Prentice Hall, 2004.
[Lewis, 1951] E. B. Lewis. Pseudoallelism and gene evolution. *Cold Spring Harbor Symposia on Quantitative Biology*, 16: 159–172, 1951.
[Lewis, 1978] E. B. Lewis. A gene complex controlling segmentation in *Drosophila*. *Nature*, 276: 565–570, 1978.
[Lewontin, 1974] R. C. Lewontin. The analysis of variance and the analysis of causes. *American Journal of Human Genetics, 26*(3), 400-411, 1974.
[Lewontin, 1991] R. C. Lewontin. Twenty-five years ago in *Genetics*: Electrophoresis in the development of evolutionary genetics: Milestone or millstone? *Genetics*, 128: 657–662, 1991.

[Lewontin, 2000] R. Lewontin. *The Triple Helix: Genes, Organisms, and Environment*. Cambridge, MA: Harvard University Press, 2000.

[Lewontin and Hubby, 1966] R. C. Lewontin and J. L. Hubby. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54: 595–609, 1966.

[Lewontin and Kojima, 1960] R. C. Lewontin and K.-I. Kojima. The evolutionary dynamics of complex polymorphism. *Evolution*, 14(4): 458–472, 1960.

[Lindegren, 1953] C. C. Lindegren. Gene conversion in *Saccharomyces*. *Journal of Genetics*, 51: 625–637, 1953.

[Lindsley and Grell, 1968] D. L. Lindsley and E. H. Grell,. *Genetic Variations of Drosophila melanogaster*. Washington, DC: Carnegie Institute of Washington Publication No. 627, 1968.

[Luria and Delbrück, 1943] S. E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28: 491–511, 1943.

[MacCorquodale and Meehl, 1948] K. MacCorquodale and P. E. Meehl. On distinction between hypothetical constructs and intervening variables. *Psychological Review*, 55: 95–107, 1948.

[Maniatis and Tasic, 2002] T. Maniatis and B. Tasic. Alternative pre-mRNA splicing and proteome expression in metazoans. *Nature*, 418: 236–243, 2002.

[Mattick, 2003] J. S. Mattick. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays*, 25(10): 930–939, 2003.

[McClintock, 1951] B. McClintock. Chromosome organization and genic expression. *Cold Spring Harbor Symposia on Quantitative Biology*, 16: 13–47, 1951.

[McGinnis *et al.*, 1984a] W. McGinnis, R. Garber, J. Wirz, A. Kuroiwa, and W. A. Gehring. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. *Cell*, 37: 403–408, 1984a.

[McGinnis *et al.*, 1984b] W. McGinnis, M. S. Levine, E. Hafen, A. Kurroiwa, and W. Gehring. A conserved DNA sequence in homeotic genes of the Drosophila antennapedia and bithorax complexes. *Nature*, 308(5958): 428–433, 1984b.

[Meijer, 1985] O. G. Meijer. Hugo de Vries no Mendelian? *Annals of Science*, 42: 189–232, 1985.

[Mendel, 1866/1966] G. Mendel. Versuche über Pflanzenhybriden. *Verhandlungen Naturforscher Verein, Brünn*, 4: 3–47, 1866. Translated and reprinted as 'Experiments on plant hybrids' in C. Stern and E. Sherwood (eds.), *The Origins of Genetics: A Mendel Sourcebook*, San Francisco, W. H. Freeman, 1966.

[Meselson and Stahl, 1958] M. Meselson and F. W. Stahl. The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Science, Washington*, 44: 671–682, 1958.

[Mitchell, 1955] M. B. Mitchell. Aberrant recombination of pyridoxine mutants of *Neurospora*. *Proceedings of the National Academy of Sciences, Washington*, 41: 215–220, 1955.

[Mohr, 1924] O. L. Mohr. A genetic and cytological analysis of a section deficiency involving four units of the X-chromosome in *Drosophila melanogaster*. *Zeitschrift für induktive Abstammungs- und Vererbungslehre*, 32: 108–232, 1924.

[Monaghan and Corcos, 1990] F. V. Monaghan and A. F. Corcos. The real objective of Mendel's paper. *Biology and Philosophy*, 5: 267–292, 1990.

[Monod *et al.*, 1963] J. Monod, J. P. Changeux, and F. Jacob. Allosteric proteins and cellular control systems. *Journal of Molecular Biology*, 6: 306–329, 1963.

[Morange, 2000] M. Morange. The developmental gene concept: History and limits. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, pages 193–215. Cambridge and New York: Cambridge University Press, 2000.

[L. V. Morgan, 1922] L. V. Morgan. Non-criss-cross inheritance in *Drosophila melanogaster*. *Biological Bulletin, Woods Hole*, 42: 267–274, 1922.

[Morgan, 1910a] T. H. Morgan. Chromosomes and Heredity. *The American Naturalist*, 44: 449–498, 1910a.

[Morgan, 1910b] T. H. Morgan. Sex limited inheritance in Drosophila. *Science*, 32: 120–122, 1910b.

[Morgan, 1911] T. H. Morgan. Random segregation versus coupling in Mendelian inheritance. *Science*, 34(873): 384, 1911.

[Morgan, 1913a] T. H. Morgan. Factors and unit characters in Mendelian heredity. *The American Naturalist*, 47(553): 5–15, 1913a.

[Morgan, 1913b] T. H. Morgan. *Heredity and Sex*. New York: Columbia University Press, 1913b.
[Morgan, 1934] T. H. Morgan. *Embryology and Genetics*. New York: Columbia University Press, 1934.
[Morgan, 1935] T. H. Morgan. The relation of genetics to physiology and medicine. Nobel Lecture, Stockholm, June 4, 1934. *Scientific Monthly*, 41: 5–18, 1935.
[Morgan *et al.*, 1915] T. H. Morgan, A. H. Sturtevant, H. Muller, and C. B. Bridges. *The Mechanism of Mendelian Heredity*. New York: Henry Holt and Company, 1915.
[Moss and Poethig, 2002] E. G. Moss and R. S. Poethig. MicroRNAs: Something new under the sun. *Curremt Biology*, 12: R688–R690, 2002.
[Moss, 2003] L. Moss. *What Genes Can't Do*. Cambridge: The MIT Press, 2003.
[Muller, 1914] H. J. Muller. A gene for the fourth chromosome of *Drosophila*. *The Journal of Experimental Zoology*, 17: 325–336, 1914.
[Muller, 1916] H. J. Muller. The mechanism of crossing over. *The American Naturalist*, 50(592–595): 193–221, 284–305, 351–366, 421–434, 1916.
[Muller, 1920] H. J. Muller. Are the factors of heredity arranged in a line? *The American Naturalist*, 54(631): 97–121, 1920.
[Muller, 1922] H. J. Muller. Variation due to change in the individual gene. *The American Naturalist*, 56: 32–50, 1922.
[Muller, 1927a] H. J. Muller. Artificial transmutation of the gene. *Science*, 66: 84–87, 1927a.
[Muller, 1927b] H. J. Muller. Quantitative methods in gene research. *The American Naturalist*, 61: 407–419, 1927b.
[Muller, 1928] H. J. Muller. The measurement of gene mutation rate in Drosophila, its high variability, and its dependence upon temperature. *Genetics*, 13: 279–357, 1928.
[Muller, 1929] H. J. Muller. The gene as the basis of life. *Proceedings of the International Congress of Plant Sciences*, 1: 897–921, 1929.
[Muller, 1935] H. J. Muller. On the incomplete dominance of the normal allelomorphs of white in *Drosophila*. *Journal of Genetics*, 30: 407–414, 1935.
[Muller, 1938] H. J. Muller. The position effect as evidence of the localization of the immediate products of gene activity, *Proceedings of the XVth International Physiological Congress Leningrad-Moscow, August 9th to 16th 1935*, pages 587–589, State Biological and Medical Press, 1938.
[Muller, 1940] H. J. Muller. An analysis of the process of structural change in chromosoms of *Drosophila*. *Journal of Genetics*, 40: 1–66, 1940.
[Muller, 1947] H. J. Muller. The gene. *Proceedings of the Royal Society, London, series B*, 134: 1–37, 1947.
[Muller, 1950] H. J. Muller. Evidence of the precision of genetic adaptation. *The Harvey Lectures*, 43: 165–229, 1950.
[Muller, 1956] H. J. Muller. On the relation between chromosome changes and gene mutations. *Brookhaven Symposia in Biology*, 8(591): 126–147, 1956.
[Muller and Herskowitz, 1954] H. J. Muller and I. H. Herskowitz. Concerning the healing of chromosome ends produced by breakage in *Drosophila melanogasrer*. *The American Naturalist*, 88: 177–208, 1954.
[Nilsson-Ehle, 1909] H. Nilsson-Ehle. Kreuzungsuntersuchungen an Hafer und Weizen. *Lunds Universität Aarskrift N. F. Afd., Ser. 2*, 5(2): 1–122, 1909.
[Nirenberg and Matthaei, 1961] M. W. Nirenberg and J. H. Matthaei. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Science, Washington*, 47(10): 1588–1602, 1961.
[C. E. Novitski, 2004] C. E. Novitski. Revision of Fisher's analysis of Mendel's garden peas experiments. *Genetics*, 166(3): 1139–1140, 2004.
[E. Novitski, 2004] E. Novitski. On Fisher's criticism of Mendel's results with the garden pea. *Genetics*, 166(3): 1133–1136, 2004.
[Nüsslein-Volhard and Wieschaus, 1980] C. Nüsslein-Volhard and E. Wieschaus. Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287: 795–801, 1980.
[Ohno, 1970] S. Ohno. *Evolution by Gene Duplication*. Berlin: Springer, 1970.
[Olby, 1974] R. C. Olby. *The Path to the Double Helix: The Discovery of DNA*. London: Macmillan, 1974.
[Olby, 1979] R. Olby. Mendel no Mendelian? *History of Science*, 17: 53–72, 1979.
[Olby, 1985] R. Olby. *Origins of Mendelism* (2nd ed.). Chicago: University of Chicago Press, 1985.

[Olby, 1987] R. Olby. William Bateson's introduction of Mendelism to England: A reassessment. *British Journal for the History Science*, 20: 399–420, 1987.
[Orel, 1984] V. Orel. *Mendel* (S. Finn, Trans.). Oxford: Oxford University Press, 1984.
[Orel, 1996] V. Orel. *Gregor Mendel: The First Geneticist* (S. Finn, Trans.). Oxford New-York Tokyo: Oxford University Press, 1996.
[Orgel and Crick, 1980] L. E. Orgel and F. H. C. Crick. Selfish DNA: The ultimate parasite. *Nature*, 284: 604–607, 1980.
[Oyama, 1985] S. Oyama. *The Ontogeny of Information. Develpmental Systems and Evolution*. Cambridge: Cambridge University Press, 1985.
[Oyama *et al.*, 2001] S. Oyama, P. E. Griffiths, and R. D. Gray (eds.). *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge: MIT Press, 2001.
[Painter, 1934] T. S. Painter. A new method for the study of chromosome aberrations and the plotting of chromosome maps in Drosophila melanogaster. *Genetics*, 19: 175–188, 1934.
[Panshin, 1935] I. B. Panshin. The analysis of a bilateral mosaic mutation in *Drosophila melanogaster* (Translated from Russian). *Akademiya Nauk SSSR. Trudy Instituta Genetiki*, 10: 227–232, 1935.
[Panshin, 1936] I. B. Panshin. A demonstration of the specific nature of position effect. *Compts Rendus (Doklady) Academie Science U.R.S.S., N.S.*, 1(10): 83–86, 1936.
[Pardue and Gall, 1970] M. L. Pardue and J. G. Gall. Chromosomal localization of mouse satellite DNA. *Science*, 168: 1356–1358, 1970.
[Perkel, 2004] J. M. Perkel. Validating the interactome, *The Scientist*, . 21 June 2004. Retrieved from the World Wide Web:
`www.the-scientist.com/yr2004/jun/feature_040621.html`
[Pontecorvo, 1958] G. Pontecorvo. *Trends in Genetic Analysis*. New York: Columbia University Press, 1958.
[Provine, 1971] W. B. Provine. *The Origins of Theoretical Population Genetics*. Chicago and London: The University of Chicago Press, 1971.
[Raffel and Muller, 1940] D. Raffel and H. J. Muller. Position effect and gene diversity considered in connection with three strikingly similar scute mutations. *Genetics*, 25: 541–583, 1940.
[Rheinberger, 1997] H.-J. Rheinberger. *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford, California: Stanford University Press, 1997.
[Rheinberger, 2000a] H.-J. Rheinberger. Ephestia: The experimental design of Alfred Kühn's physiological developmental genetics. *Journal of the History of Biology*, 33(3): 535–576, 2000a.
[Rheinberger, 2000b] H.-J. Rheinberger. Mendelian inheritance in Germany between 1900-1910: The case of Carl Correns (1864-1933). *Comptes Rendus de l'Academie des Sciences/ Life Sciencs*, 323(12): 1089–1096, 2000b.
[Rieger *et al.*, 1976] R. Rieger, A. Michaelis, and M. M. Green. *Glossary of Genetics and Cytogenetics* (4th ed.). Berlin: Springer-Verlag, 1976.
[Roll-Hansen, 1978] N. Roll-Hansen. The genotype theory of Wilhelm Johannsen and its relation to plant breeding and the study of evolution. *Centaurus*, 22: 201–235, 1978.
[Rosenberg, 1985] A. Rosenberg. *The Structure of Biological Science*. Cambridge: Cambridge University Press, 1985.
[Roux, 1894/1986] W. Roux. The problems, methods, and scope of developmental mechanics. In J. Maienschein (ed.), *Defending Biology: Lectures from the 1890s*, pages 104–148. Cambridge, MA: Harvard University Press. German original, 1894.
[Sapp, 1990] J. Sapp. *Where the Truth Lies: Franz Moewus and the Origins of Molecular Biology*. Cambridge: Cambridge University Press, 1990.
[Sarkar, 1998] S. Sarkar. *Genetics and Reductionism*. Cambridge: Cambridge University Press, 1998.
[Sarkar, 1999] S. Sarkar. From the *Reaktionsnorm* to the adaptive norm: The norm of reaction, 1909-1960. *Biology and Philosophy*, 14(2): 235–252, 1999.
[Schaffner, 1976] K. Schaffner. Reduction in biology: Prospects and problems. In R. S. Cohen, C. A. Hooker, G. Pearce, A. C. Michalos, and J. W. van Evra (eds.), *Proceedings of the 1974 Biennial Meeting of the Philosophy of Science Association*, pages 613–632. Dordrecht: Reidel, 1976.
[Schrödinger, 1962] E. Schrödinger. *What Is Life? The Physical Aspect of the Living Cell*. Cambridge: University Press, 1962. First published 1944.

[Schwartz, 1998] S. Schwartz. *The significance of the trait in genetics, 1900-1945*. Unpublished Ph.D. Dissertation, The Hebrew University of Jerusalem, Jerusalem, 1998.

[Schwartz, 2000] S. Schwartz. The differential concept of the gene. In P. J. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, pages 26–39. Cambridge and New York: Cambridge University Press, 2000.

[Schwartz, 2002] S. Schwartz. Characters as units and the case of the presence and absence hypothesis. *Biology and Philosophy*, 17(3): 369–388, 2002.

[Stadler, 1928] L. J. Stadler. Mutations in barley induced by X-ray and radium. *Science*, 68: 186–187, 1928.

[Stadler, 1954] L. J. Stadler. The gene. *Science*, 120: 811–819, 1954.

[Stahl, 1961] F. Stahl. A chain model for chromosomes. *Journal de Chemie Physique*, 58: 1072–1077, 1961.

[Stahl, 1979] F. W. Stahl. *Genetic Recombination: Thinking About It in Phage and Fungi*. San Francisco: W. H. Freeman, 1979.

[Stamhuis, 1995] I. H. Stamhuis. The "rediscovery" of Mendel's laws was not important to Hugo de Vries: Evidence from his letters to Jan Willem Moll. *Folia Mendeliana*, 30: 13–30, 1995.

[Stamhuis *et al.*, 1999] I. H. Stamhuis, O. G. Meijer, and E. J. A. Zevenhuizen. Hugo de Vries on heredity, 1889-1903: Statistics, Mendelian laws, pangenes, mutations. *Isis*, 90(2): 238–267, 1999.

[Stent, 1968] G. S. Stent. That was the molecular biology that was. *Science*, 160: 390–395, 1968.

[Stent, 1970] G. S. Stent. DNA. *Daedalus*, 99: 909–937, 1970.

[Stern, 1931] C. Stern. Zytologisch-genetische Untersuchungen als Beweise für die Morgansche Theorie des Faktorenaustauschs. *Biologisches Zentralblat*, 51(10): 547–587, 1931.

[Stern, 1943] C. Stern. The Hardy-Weinberg law. *Science*, 97(2510): 137–138, 1943.

[Stern, 1955] C. Stern. Two or three bristles. *Science in Progress*, 9: 41–84; 327–328, 1955.

[Stern and Sherwood, 1966] C. Stern and E. Sherwood (eds.). *The Origin of Genetics: A Mendel Sourcebook*. San Francisco: W. H. Freeman, 1966.

[Streisinger *et al.*, 1967] G. Streisinger, Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. Frameshift mutations and the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology*, 31: 77–84, 1967.

[Sturtevant, 1913] A. H. Sturtevant. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology*, 14: 43–59, 1913.

[Sturtevant, 1929] A. H. Sturtevant. The claret mutant type of *Drosophila simlans*: a study of chromosome elimination and cell-lineage. *Zeitschrift für Wissenschaftliche Zoology*, 135: 323–356, 1929.

[Sulston and Ferry, 2002] J. Sulston and G. Ferry. *The common Thread: A Story of Science, Politics, Ethics and the Human Genome*: The Joseph Henry Press, 2002.

[Sutton, 1903] W. Sutton. The chromosomes in heredity. *Biological Bulletin*, 4: 231–251, 1903.

[Tabery, 2004] J. G. Tabery. The "evolutionary synthesis" of George Udny Yule. *Journal of the History of Biology*, 37(1): 73–101, 2004.

[Theunissen, 1994] B. Theunissen. Closing the door on Hugo de Vries' Mendelism. *Annals of Science*, 51: 225–248, 1994.

[Timoféeff-Ressovsky *et al.*, 1935] N. W. Timoféeff-Ressovsky, E. G. Zimmer, and M. Delbrück. Über die Natur der Genmutation und der Genstruktur. *Nachrichten der Biologischen Gesellschaft für Wissenschaft, Göttingen*, 1: 189–245, 1935.

[Troland, 1917] L. T. Troland. Biological enigmas and the theory of enzyme action. *The American Naturalist*, 51(606): 321–350, 1917.

[Vogt, 1926] O. Vogt. Psychiatrisch wichtige Tatsachen der zoologisch-botanischen Systematik. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 101: 805–832, 1926.

[Waddington, 1957] C. H. Waddington. *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology*. London: George Allen & Unwin, 1957.

[Watson and Crick, 1953a] J. D. Watson and F. H. C. Crick. Genetical implications of the structure of deoxyribose nucleic acid. *Nature*, 171: 964–967, 1953a.

[Watson and Crick, 1953b] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids. *Nature*, 171: 737–738, 1953b.

[Westergaard, 1958] M. Westergaard. The mechanism of sex determination in dioecious flowering plants. *Advances in Genetics, 9*, 217-281, 1958.

[Whitehouse, 1965] H. L. K. Whitehouse. *Towards an Understanding of the Mechanism of Heredity*. London: Edward Arnold, 1965.
[Wilson, 1893/1986] E. B. Wilson. The mosaic theory of development. In J. Maienschein (ed.), *Definding Biology: Lectures from the 1890s*, pages 67–80. Cambridge, MA: Harvard University Press, 1986. First published, 1893.
[Wilson, 1896] E. B. Wilson. *The Cell in Development and Inheritance*. New York: Macmillan, 1896.
[Winkler, 1930] H. Winkler. *Die Konversion der Gene*. Jena: Fischer, 1930.
[Woltereck, 1909] R. Woltereck. Weitere experimentelle Untersuchungen über Artveränderung, speziell über des Wesen quantitativer Artunterschiede bei Daphnien. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, 19: 110–173, 1909.
[Wright, 1986] S. Wright. *Evolution, Selected Papers*, W. B. Provine, ed. Chicago: The University of Chicago Press, 1986.
[Yanofsky *et al.*, 1964] C. Yanofsky, B. C. Carlton, J. R. Guest, D. R. Helinski, and U. Henning. On the colinearity of gene structure and protein structure. *Proceedings of the National Academy of Science, Washington*, 51(2): 266–272, 1964.
[Yule, 1902] U. Yule. Mendel's laws and their probable relations to intra-racial heredity. *The New Phytologist*, 1(9 & 10): 193–207 & 222–238, 1902.
[Zuckerman and Lederberg, 1986] H. Zuckerman and J. Lederberg. Postmature scientific discovery? *Nature*, 324(6098): 629–631, 1986.

THE DEVELOPMENT OF POPULATION GENETICS

Margaret Morrison

Genetics, like many other sciences that underwent rapid and extensive development in the twentieth century, now encompasses a broad range of topics and fields. Instead of genetics being considered as a branch of biology it is now a full fledged science in its own right. Although typically concerned with the problems of heredity and variation, genetics has developed in primarily six different directions, the oldest being *experimental breeding* which dates back hundreds of years and involves the study of several generations of plants and animals. It relies primarily on statistical methods as its standard research methodology. *Pedigree analysis*, which shows the inheritance of specific traits traced along members of a family line, is frequently used in cases where experimental breeding isn't practical. *Cytology* studies the chromosomes and cellular infrastructure that are the keys to heredity while *biochemistry* (specifically molecular biology) investigates the nature of genes themselves, how they function and reproduce. An analysis of gene action allows us to chart the course of amino acid breakdown in cells and determine how abnormalities arise when a gene fails to produce a particular enzyme. *Population genetics* deals with the distribution of genes in different populations and traces population migration and mixing of races by analysing the frequency of various blood antigens. These same methods allow for the analysis of fossil genetic material used for tracing the process of heredity over thousands of years. Finally, *genetic recombination* provides the foundation for extensive and detailed maps of the nucleotide sequences of gene molecules which form the basis of the human genome project.

Just as in history and philosophy of physics where debates about who "discovered" a particular particle or who first used a particular term are part of the landscape, so too in discussions of genetics; debates about its discovery, who first used the phrase and whether Mendel was really the founder of what came to be called Mendelian genetics are not uncommon. These are not the kinds of issues I will be concerned with here. Instead my focus is on the origins and development of population genetics up to 1930; a story that deals specifically with the fusion of Darwinian natural selection with Mendelian genetics into a theory that could explain what Darwinism couldn't — variation in a population. The development of population genetics is also important because it was able to provide knowledge about heritable traits that were simply outside the grasp of the experimental breeder. Its success was in due, in part, to sophisticated statistical/mathematical

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

techniques developed by the early population geneticists, especially R. A. Fisher (1890-1962). In that sense then one needs to understand the history of population genetics as coincident with another developing science of the early twentieth century — statistics.

Genes are the physical units of heredity and in modern terms are typically defined as a segment of genetic material that determines the sequence of amino acids in specific polypeptides. Gregor Mendel's work on crossbreeding pea plants [1865] is usually referred to as the beginning of genetics proper since it was in that context that his basic principles of inheritance were specified. Leaving aside questions of what exactly Mendel discovered and what constitutes "Mendelism", it is nevertheless important to mention some of the basic features of the story. What Mendel set forth in his 1865 paper was essentially a law of serial development based on straightforward combinatory analysis which stated that a hybrid of a particular character would produce an offspring distributed according to specific proportions. Based on his experiments with garden peas he analysed the numerical data for pure parental forms A and a and the hybrid Aa and found that the first generation from hybrid were distributed according to the ratio 1A:2Aa:1a. This analysis was then extended to several independent characters or traits. He went on in a later part of the paper entitled 'the reproductive cells of hybrids' to explain the combination of these traits as the result of the union of reproductive cells, each of which carries a particular trait. In that sense there was a correspondence between the combination series of the traits and the union of the gametes during fertilization. What this means is that if the individual is a hybrid then the two characters or *anlagen* will act separately. Although Mendel himself didn't specify the two independent laws with which he is frequently associated (segregation, which states that a cross should be defined in terms of a pair of antagonistic characters and independent assortment which claims that characters assort independently) the conclusions on which they are based can clearly be found in his work. And, while he used the terms dominant and recessive he did not formulate a specific law of dominance as such.

For thirty-five years Mendel's work went largely unnoticed, then in 1900 it was "rediscovered" independently by Hugo De Vries (1848–1935), a Dutch botanist, Carl Correns (1864–1933), a German biologist, and E. von Tschermak (1871–1962), an Austrian naturalist. With this rediscovery came new interpretations and extensions of Mendel's original results, some of which will be discussed below. However, even before 1900 other work was being done on the notion of heredity and its physiological and statistical bases. In order to appreciate the implications of these early accounts of heredity and the notion of discontinuous variation for the theory of human evolutionary development we need to go back to 1865, but to a different context — the work of Francis Galton (1822-1911). Galton was the first cousin of Charles Darwin and studied medicine at Cambridge. He abandoned a medical career to explore Africa and in 1853 received a gold medal from Royal Geographical Society for his achievements. In the early 1860s he did extensive work with weather maps, using graphical methods for the analysis of multivariate

data. From 1865 onwards his main focus was study of heredity. Galton's work is especially important because he was the first to apply statistical methods to the study of heredity which in turn led to the growth of biometry, a methodology that would eventually be used by Fisher in developing the mathematical foundations of population genetics.

1 EARLY DAYS: THE LAW OF ANCESTRAL HEREDITY

In order to appreciate Galton's contributions we need to know a few details about the historical context in which he was working. In the *Origin of Species* [1859] Darwin claimed that species became adapted to their environment through natural selection acting on heritable variation. One of the fundamental difficulties with the Darwinian theory of evolution was that it lacked any developed notion of heredity that could serve as the basis for the operation of selection. Instead the Darwinians simply assumed that the characteristics of the offspring resulted from a blending of the characteristics of the parents. Evolution by natural selection was a continuous and gradual process. But this gave rise to further difficulties: It was unable to account for "sports" (large variations that occurred in single instances); it could not explain the increase or decrease in a species' store of variations, and it provided no way of determining whether particular characteristics in an individual were due to heredity or the environment. In other words, the laws governing evolution could not be given until those governing variation were also known. This was part of the problem Galton attempted to tackle, and to that extent it was continuous with the problem that occupied the later Mendelians. Unfortunately a solution to the problems of heredity transmission yielded a further problem for Darwinism. The notion of discontinuous variation which was crucial to Galton's theory of heredity, and later, Mendelism, was in direct opposition to the continuous variation that seemed crucial to Darwinism. So, could one reconcile an explanation of the mechanisms of heredity that was seemingly at odds with fundamental aspects of evolution by natural selection? It was that issue that later gave rise to population genetics and the synthesis of Darwinain selection with Mendelian genetics.

Galton's first paper entitled "Hereditary Talent and Character" [1865] attempted to show that mental and moral characters of humans are the product of the combined action of natural selection and heredity. The paper outlined a general concept of heredity that involved the continuity of the "germ line" and the need to explain both resemblance and variability within families and races in general, ideas that Galton went on to develop in subsequent work over the next thirty five years. The paper also provided an exposition of what became known as the law of ancestral heredity which he understood as supporting his hypothesis of discontinuous evolution. Based on no obvious data except lists of able men who had able relatives, the law stated that: "The share a man retains in the constitution of his remote descendants is inconceivably small. The father transmits, on average, one-half of his nature, the grandfather one-fourth, the great-grandfather one-eight; the share

decreasing step-by-step in a geometrical ratio with great rapidity".[1] This also marked the introduction of statistical methods into biology, the defining feature of what would become, in the hands of Karl Pearson, the science of biometry.

In 1869 with the publication of *Hereditary Genius* Galton elaborated his idea that evolution proceeded by discontinuous leaps but he needed a theory of heredity to back it up. By 1875 he developed the view that hereditary qualities were concentrated in the reproductive organs rather than embedded in gemmules, which according to Darwin's pangenesis theory were formed in all parts of an organism. The germ plasm or "stirp", as Galton called it, was continuously inherited by each generation with very little alteration. Variations were caused by alterations of this germ plasm and were distinct. The publication of *Finger Prints* in 1892 marked the forceful declaration that evolution proceeded by jerks through successive sports, each of which is favoured by natural selection. Prior to this however, in the period between 1877 and 1889 Galton gathered data on the inheritance of size in sweet peas as well as the stature, eye colour, temper, disease and artistic ability in man, all in an attempt to derive a quantative law of regression, which was given a complete formulation in *Natural Inheritance* [1889]: "... the deviation of the Sons from P [the median stature of the general population] are, on average, equal to one-third of the deviation of the Parent from O, and in the same direction... If P+(+- D) be the stature of the Parent , the Stature of the offspring will on the average be P+(+-1/3 D).[2]

Although the law was based on data about stature he thought it applicable to non-blending inheritance like eye colour in humans and coat colour in Bassett hounds where the total heritage was represented by percentages in the offspring. In other words, since coat colour fell into one of two categories (tricolour or lemon and white) one could accurately predict that one quarter of the litter would take after the mother, one quarter after the father, one eighth after each grandparent, a sixteenth after each great grandparent etc. Indeed his work on regression and the law of ancestral heredity convinced him of the ineffectiveness of selection on individual differences, a view first put forward in *Heredity Genius* [1869] where he claimed that "because an equilibrium between deviation and regression will soon be reached, ... the best of the offspring will cease to be better than their own

[1]Pearson pointed out that Galton should have referred to mid-parent, mid-grandparent etc., something that he did do in a later formulation [1885]. The mid-parent is a fictitious individual whose height, for example, was equal to half the sum of the paternal and adjusted maternal heights.

[2]Although the data suggested that the average regression of mid-filial stature upon mid-parental stature was around 3/5, Galton substituted 2/3 because it was simpler. As Provine [1971] points out, the value 2/3 enables him to calculate the mid-filial regression on a single parent without sufficient data. If each parent contributes equally the value will be $\frac{1}{2}$ of the joint contribution or $\frac{1}{2}$ of that of the mid-parent. Hence, the average regression from the parental to the mid-filial stature will be $\frac{1}{2}$ of 2/3 which equals 1/3. His linear regression equation took the following form: x=(2/3)y, were x=X(filial stature)-$M_{f,}$(mean filial stature), y=Y(midparental stature)-M_p(mean midparental stature). 2/3 is called the regression coefficient and where x and y have the same variance it is identical with the correlation coefficient, the value of which is given by x/y.

sires and dams." (p.34). In other words, any extremes caused by selection would be returned to the centre by regression; hence, "sports", which he considered quite stable, must be the only effective source of evolutionary variation. This of course was in direct conflict with Darwinism, which saw "sports" as having no role in evolution due to their obliteration by blending inheritance.

Galton's arguments against blending inheritance, which were backed up by powerful statistical analyses of the data turned out to be incorrect. And, what initially looked like a clever and potentially devastating strategy against the Darwinians was actually turned to their advantage by Karl Pearson (1857–1936). Pearson graduated with mathematical honours from Cambridge in 1879 but before embarking on his work in statistics he was a scholar of German history, folklore and philosophy. In addition, he worked in the philosophy of science, the theory of elasticity and was active in politics at the University of London, especially on issues regarding socialism and the education of women. With the assistance of his colleague the marine biologist W.F.R. Weldon 1860–1906), Pearson extended Galton's statistical work into biology while his Mendelian views were developed in Britain primarily by the biologist WilliamBateson (1861–1926).[3] Although Weldon was not a mathematician he was particularly interested in Galton's methods but required the help of the more able Pearson to fully apply them in the context of animal evolution. Bateson was interested in another aspect of Galton's work, the discontinuously varying traits such as white and red flower colour. This issue was discussed by Galton in *Natural Inheritance* where he put forward his views about evolution involving a discontinuous process. Pearson saw Galton's work not as "a biological hypothesis, but the mathematical expression of statistical variates... [which] can be applied... to many biological hypotheses" [Pearson, 1930, 21]. His goal was to give the work a more rigorous foundation and hence to secure a scientific foundation for natural selection, a task which involved an evaluation of the heredity law, specifically the regression constants in the geometric series (1/2, 1/4...).

In a paper in 1895 Pearson pointed out that the multiple regression coefficients 1/4, 1/16, 1/64 etc. were incompatible with the assumption that correlations of the offspring with the individual parent, grandparent etc. form the series r, r^2, r^3, etc. He showed that any such series caused all the coefficients except the first or parental coefficient to vanish and reduced the ancestral multiple regression to a simple biparental inheritance. Hence, the parental characters would determine completely those of the offspring and contrary to what Galton had assumed, after the relaxation of selection and commencement of inbreeding, there would be no further regression after the first reproduction.[4] In 1895 Pearson formulated a version of the ancestral law showing its connection to the law of regression. This new law took the form of a multiple regression equation of offspring on mid-midparental ancestry:

[3]Although it is sometimes suggested that Bateson's account of Mendelism deviated from Mendel's own formulation but that is not a topic I want to explore here.

[4]See also [Pearson, 1930, 39].

$$P_0 = \tfrac{1}{2}\ (\sigma_0/\sigma_1\ P_1) + \tfrac{1}{4}(\sigma_0/\sigma_2\ P_2) + 1/8\ (\sigma_0/\sigma_3\ P_3)\ldots$$

P_0 is the predicted deviation of an offspring from the generation mean, P_1 is a linear function of the deviation of the mid-parent from that generation mean, P_2 similarly for the mid-grandparent and σ_0, $\sigma_1\ldots$ the standard deviation of the appropriate generations of the offspring. From this formula Pearson derived theoretical values for various regression and correlation coefficients between relatives.[5]

Basically, Galton's shortcomings were traceable to a lack of knowledge of the correct formula for multiple regression.[6] Had he been able to get beyond regression in two variates the contradiction between his law of ancestral heredity and his interpretation of regression would have been clear; that is, the connection between his mid-parental heredity and his law of ancestral heredity requires a knowledge of multiple correlation. And, it was this interpretation of regression that was responsible for his supporting discontinuous evolution. With these corrections in place Pearson [1898] was able to reconcile Galton's work with his own views about variation, claiming that the ancestral law formed the fundamental principle of heredity from which all the numerical data of inheritance could be deduced, at least to a first approximation. The extent to which Pearson distanced the law from specific biological hypotheses is evident from his remark that the confidence he put in the truth of the law was "not measured" by any of Galton's empirical research however strong that evidence may be; instead his support for it was grounded in the fact that it gave "*a priori* the correlation between parents and offspring and that this correlation is practically identical with the value [he had] determined from these and other observations" [1895, 387].[7]

There are a number of different issues intertwined here, Mendelism vs. blending inheritance, Darwinism vs. saltation and biology vs. statistics, each of which needs to be clarified before proceeding. As we saw above Galton's work engaged both Mendelism and Darwinism in that he advocated natural selection as an important feature of evolutionary change, but his theory of heredity was more in line with what would later be identified as Mendelism than with the gradualism that characteristic of Darwinism. The Darwinian account emphasised the action of selection on small alterations that resulted from a blending of the ancestral traits while Mendelism advocated discontinuously varying traits. This distinction between continuity and discontinuity was also at the source of the difference between Darwinism and saltation where the latter describes the evolutionary process as involving sudden changes in the equilibrium of a species where there is a leap

[5]He also generalized the geometric series of partial regression coefficients which raised the parental correlations.

[6]See [Pearson, 1930, 77–78].

[7]Pearson further refined Galton's law in 1900 where he distinguished a law of reversion from the true ancestral law. In cases of blended inheritance the law of ancestral heredity predicts the probable character of the individual produced by a given ancestry while in cases of alternative inheritance the law tells us the percentages of the total offspring which, on average, revert to each ancestral type. The first is the true law of ancestral heredity while the second he termed the law of reversion. Both laws assume a geometrically waning ancestral dilution.

from one position of organic stability to another. Proponents of saltation, particularly Galton, claimed that evolution could never progress by small continuous changes of the sort characterised by Darwinism because there would be continual regression to the mean. These different views about the nature of heredity and evolutionary progress are, quite clearly, biological hypotheses that can be distinguished from the issue of statistical data analysis introduced by Galton and further developed by Pearson and Weldon. The latter is concerned with devising methods for predicting how traits will be distributed in future generations, something that needn't be associated with any specific biological theory of heredity.

2 STATISTICAL DARWINISM VS. MENDELIAN FACTORS

This distancing between biology and statistics is further evident in the work of Weldon, Pearson's colleague and collaborator. Although Weldon's approach initially emphasised morphological studies he was a strong proponent of the use of statistical methods, and claimed that the problem of animal evolution was essentially a statistical problem. Pearson was aware that such analysis required a more advanced theory of statistics than was currently available, something he went on to develop in a series of papers written over the next decade. It was this application of statistical tools designed to measure variability and correlation and the ways in which these influenced various kinds of selection that defined the new science of biometry. In that sense then the biometrician's goal was to develop a purely statistical theory of evolution devoid of the physiological aspects of heredity that interested Galton. Like true empiricists the biometrists rejected causal notions and instead spoke of "coefficients" that would empirically measure heredity based on such factors as environment, homogomy and natural selection itself. Pearson's goal was to redefine the problem of selection as a problem about heredity which he could then treat from within the theory of correlation. In the hands of Pearson heredity was nothing other than a correlation coefficient — its true measure was the numerical correlation between some characteristic or organ as it occurs respectively in parent and offspring. It was similar to a force whose underling causal mechanism was deemed irrelevant.

Even before the rediscovery of Mendel's work in 1900, Bateson, who became the spokesman for Mendelism in Britain, published *Materials for the Study of Variation* [1894]. The book was a sustained attack on the Darwinian emphasis on continuity and its reliance on adaptation as the sole agent responsible for directing evolution. Bateson argued that evolution should be investigated by uncovering the process by which new characters were produced by variation. Through his studies on hybridization he attempted to show how varieties could be distinguished by the behaviour of unit characters. Although he was unaware of Mendelian ratios at the time, his later 1902 book, *Mendel's Principles of Heredity*, contained a translation of Mendel's paper and a strong defense of his laws. In that same year he also introduced the term "allelomorph" to define the alternative characters in segregating pairs and used the terms "heterozygote" and "homozygote" to replace

"hybrid" and "pure type".

During the period 1889-1900 de Vries was also promoting the idea that hereditary qualities were independent units and argued that the species, as the unit of heredity transmission, should be replaced by unit characters. With the rediscovery of Mendel's work in 1900 came several reconceptualizations of it, including disagreements between de Vries and Correns about the extent to which the law of segregation was applicable. The various details of how Mendel's work gradually gave rise to what Bateson termed the science of "genetics" in 1905 are too numerous to document here. However, the main difference centred on the idea that each hereditary character had a corresponding structure in each individual. In Mendel's work there was no such notion; that is to say, no theory that inherited characters were determined by a finite number of units present in two copies in each individual, with each one being able to take two alternative forms (alleles). The identification of the cytological structure of the nucleus with the structure of the Mendelian crosses was established in 1902 by Walter Sutton (1877-1916) and along with it came the reinterpretation of Mendel's schema in terms of physical division and combination. This became evident by the replacement of Mendel's original representation of characters as A, Aa and a, which indicated the appearance of a trait, with AA, Aa and aa, indicating the germ structure.

In 1902 while the Mendelians were developing and extending the theory of inheritance Pearson engaged in a debate with Bateson about the *methodological* superiority of biometry over Mendelism. Their disagreement about the status of Mendelism had passed beyond the confines of biology and instead focussed on the correct techniques for assessing claims about evolutionary development. To some extent this is not surprising given what we have seen above — Pearson's insistence that the problem of evolution is a problem of statistics. But at this point in the story there seems to be a more powerful strategy emerging. Recall that Pearson used statistical techniques to argue against Galton's claim about discontinuous evolution by showing that it involved a misinterpretation of the notion of regression. In his debate with Bateson, Pearson refused to accept any biological arguments or evidence not grounded in biometrical techniques; in other words, biology, based on experiment (other than the collection of sample data) had in some sense ceased to become a source of scientific knowledge for Pearson. In fact, he [1902] went so far as to claim that one could not even have a coherent notion of discontinuous variation without the employment of biometric methods. And, because biometric methods gave the kinds of phenotypic relations that could be observed in nature, the burden of proof was on the Mendelians to show that a different distribution of variation was the correct one. But, it wasn't simply that Biometry supported the status quo. Many of the phenomena that Bateson claimed could not be biologically detected, such as differentiation between like organs and between brothers, Pearson claimed *could* be detected (where it had any sensible influence) by biometric methods. Moreover, there was *no other way* of differentiating any kind of variation from normal variation. Without dealing with the vital statistics of large populations it would be impossible to make any

progress in the theory of evolution since no tabulation of individual instances could possibly lead to definite conclusions.

However, what the new emphasis on Mendelian germ structure implied was the rejection of the "descent with modification' account of heredity that characterised Darwinism. In this latter context ancestry was a crucial feature of heredity and the biometricians went some way toward providing a quantitative analysis of the resemblance of individuals based on common ancestry. The law of ancestral heredity (in the form of a multiple regression equation) stated the contribution from each ancestor to the constitution of the individual. No mechanisms for explaining heredity were postulated, nor were they desirable. It could be completely understood in terms of descent or lineage and correlation coefficients. In contrast to this Mendelism focussed on the genetic structure of the individual, particularly the composition of the zygote with the structure of the parents uniquely determining the structure of the offspring. Once this was established ancestry was of no importance. And, as R.A. Fisher later pointed out, the methods for predicting the appearance of certain traits were fundamentally different.

With its emphasis on statistical analysis of large populations the biometricians/Darwinians could fairy accurately predict the ways in which populations would evolve. Because Mendelism used a different kind of analysis it could accurately predict the offspring of particular parents but not of populations as a whole. However, if Mendelism was to provide the mechanism of heredity that was absent in Darwin's theory then any fusion of these two would require an account of how the structure of Mendelian heredity can be analysed for populations. In other words, because selection acts on populations and not individuals we need an account of how to reconcile the individualistic nature of Mendelism with the more broadly based theory of selection. In addition we need a way of understanding, from the point of view of selection, how genetic structure changes; that is, how to account for mutation. Answers to these questions emerged in the period between 1918 and 1930 in the work of R.A. Fisher, J.B.S. Haldane (1892–1964) and Sewall Wright (1889–1988). But, before this there were developments that would prove crucial for Mendelism, especially the formulation of the Hardy-Weinberg law.

In 1902 George Udny Yule (1871–1951) showed that two populations A and a, that are the same size will reproduce the same population structure in successive generations (1A:2Aa:1a) provided there is random mating. Prior to this Yule had studied engineering at University College and Physics under Hertz in Bonn. He accepted a position as Pearson's assistant in his lab at UCL 1893 and there began his work in statistics. In the 1902 paper Yule assumed the total dominance of one character which resulted in the population remaining in equilibrium with a ratio of 3:1 dominants: recessives. A similar result, but without the assumption of dominance, was produced by Pearson in 1904. But, both Pearson and Yule considered only two Mendelian factors present in equal frequencies. It was the geneticist R.C. Punnett, sceptical of the dominance assumption, who in 1908 asked Godfrey Hardy (1877-1947), an English mathematician, to investigate the problem. He showed that regardless of the initial frequencies of the forms AA, Aa, aa, a

stable equilibrium should be established from the second generation. In other words, no dominant character would have influence over a population. The law states that the Mendelian 1:2:1 ratio will be stable from one generation to the next under conditions of random breeding, provided that there are no outside influences like mutation or selection (except for balanced rates) and the population is large enough to rule out sampling errors. This can be expressed as the expansion of the algebraic binomial $p^2 + 2pq + q^2 = 1.0$ where p is the frequency of A and q the frequency of a in the population as a whole.

The Hardy-Weinberg law was also formulated independently in 1908 by Wilhelm Weinberg (1862-1937) a German obstetrician. The law can be understood as a direct consequence of Mendel's law of segregation. Mendel's studies indicated that an immediate consequence of segregation in self-fertilizing organisms like peas was an orderly and predictable decrease in the proportion of heterozygotes in successive generations produced from a cross between two initial homozygotes. Hence, inbreeding should result in a gradual reversion of offspring populations to the original homozygotic parental types. So, when non-random mating is the rule the frequency of heterozygotes will be expected to decline; in other words, the splitting of Aa hybrids into A+2Aa+a offspring leads to a progressive reduction in the proportion of Aa hybrids. After n generations the population will have the following structure: $(2^n\text{-}1)$A:2Aa$(2^n\text{-}1)$a.

What is significant about the Hardy-Weinberg law is not so much the binomial form of the genotype frequency and the prediction of genotypes based on the stability of the population, but rather what the stability actually shows or presupposes. Despite the idealizing assumptions the stability allows us to understand something about Mendelian populations that is significant for understanding heredity and variation. What it says is that if no external forces act then there is no intrinsic tendency for the variation caused by the three different genotypes that exist in a population to disappear. It also shows that because the distribution of genotype frequencies is independent of dominance, dominance alone cannot change genotype frequencies. In other words, there is no evidence that a dominant character will show a tendency to spread or a recessive one die out. Instead the genotypes frequencies are maintained in constant proportions. The probabilistic genetic structure is conserved indefinitely; but should it be influenced by an outside force, e.g. mutation, the effect would be preserved in a new stable distribution in the succeeding generation. In other words, it expresses the variation preserving character of the hereditary mechanism and gives a specification of the number of generations required to alter gene frequencies.

This was crucial for understanding the debate between the Darwinians who advocated blending inheritance and the Mendelians. Under blending inheritance variation was thought to decrease rapidly with each successive generation, but Hardy-Weinberg shows that under a Mendelian scheme it is maintained. This pointed to yet another fundamental aspect of Mendelism, namely the discontinuous nature of the gene, and why it was important for the preservation of variation required for selection. How was it possible for the genetic structure to be main-

tained over successive generations? The reason for the stability could be traced directly to the absence of fusion which was indicative of a type of genetic structure that could conserve modification. The constancy of the Mendelian ratios can be seen as a property of genetic structure. In that sense then one can see the Hardy-Weinberg law as the beginnings of a new understanding of the role of mutation and selection and how they affect our understanding of evolution.

In 1903 the American geneticist William Castle (1867-1962) was the first to attempt an account of selection using Mendelism. His work went largely unnoticed but a more sophisticated version was given by the biologist Punnett in 1915 in a book entitled *Mimicry in Butterflies*. However, it was not until 1918 and the publication of R.A. Fisher's paper "On the Correlation between Relatives on the Supposition of Mendelian Inheritance" that a truly quantative account of the operation of selection as a process of gene replacement in Mendelian populations was given. But, establishing the relation between Mendelism and Darwinism was not an easy task; efforts to do so suffered strong opposition from Pearson's biometric school, despite the fact that Fisher himself used biometric methods in arriving at his conclusions. The way the story unfolds is important not only from the point of view of mathematising genetics but also because of the nature of the idealisations introduced by Fisher. We saw above that the Hardy-Weinberg law introduced certain idealizing assumptions about the size of populations and the fact that they were not acted upon by outside influences. What Fisher's work does is introduce other kinds of idealizing assumptions that characterise populations; assumptions that Pearson and some Mendelians felt were unacceptable for representing biological individuals and populations.

3 FROM STATISTICAL BIOLOGY TO MATHEMATICAL POPULATIONS

While Fisher is usually credited with showing that natural selection operates in Mendelian populations, thereby reconciling Darwinain and Mendelian views on heredity and evolution, it is important to emphasise that he also reconciled Mendelian and biometric *approaches* to the problem of heredity. In other words, not only did his synthesis involve a substantial claim about the mechanisms of inheritance but also a methodological solution to a problem that for Pearson was irreconcilable. The idea that Mendelism and Darwinism might be compatible was pursued and developed by Pearson in [1904]. There he showed that if a trait like stature depended linearly and additively on n independent Mendelian loci, each with complete dominance, the expectation of the distribution of stature for large n would be very close to the normal (due to the normal approximation for the binomial distribution). Although these same suppositions led to an expectation of linear regression between relatives, the regression coefficients took values that differed from Pearson's own investigations — 1/3 as opposed to 0.45-0.5. Moreover, the size of n was found not to affect the predicted value, whereas his observations led to regression coefficients that varied from organ to organ and species to species.

Instead of interpreting the results as evidence for a possible confirmation of

Mendelian principles using biometric techniques, Pearson claimed that a general pure gamete theory was too inelastic to cover the facts. In other words, there was no room for manipulation of existing theoretical assumptions that could result in a convergence of the regression coefficients to the same value. From the point of view of a Mendelian analysis there were simply too many details that needed to be considered. Although there was nothing inconsistent between biometry and such a general Mendelian theory, the laws required for a formulation of the latter (e.g. laws governing reproduction as a result of random mating, to name just one type) also needed to be tested against the inheritance statistics for the relevant populations. However, deducing all of these effects in the field of inheritance involved extremely complicated mathematical analysis — too complicated for the consideration of single cases. Pearson considered his pure gamete theory a "generalized Mendelian theory" for the case he had considered and claimed it would lead to the same system of inheritance developed by statistical description of observed facts. In other words, although this version of the Mendelian theory could not stand alone as providing general principles of inheritance it created no inconsistencies with the biometric account. The statistical discrepancies could be resolved by adding some additional assumptions to the theory or by introducing some real life complexities like homogamy or differential fertility into the theoretical scheme.

The advantage of biometric analysis was that it could attain a level of certainty that was impossible using Mendelian schemes. If one was interested in how characteristics were inherited and how populations changed over time and then one needed statistical techniques that would permit the measurement of changes in frequency of phenotypic characters, as well as large representative samples from which to measure population parameters. Once the observations were analysed into statistically defined populations they formed a "model" of the real population. Biometrical statistics was designed to apply to large samples drawn from human, plant and animal populations. And, if the samples were large enough then one could supposedly substitute the sample statistics for the population parameters (e.g. mean, standard deviation, correlation coefficient etc.). This kind of statistical practice together with the notion that species needed to be defined in population terms separated Weldon and Pearson from most biologists (not just Mendelians). Indeed, it was this concept of a "model population" that marked the differences between Pearson and Fisher, and it was Fisher's reconstruction of the notion of a population that ultimately facilitated the synthesis of Mendelism and Biometry.

Fisher studied physics and mathematics at Cambridge and had a keen interest in Eugenics. As early as 1911, during his second year at Cambridge he had familiarized himself with Pearson's work and the issues surrounding multifactorial Mendelian models. Although he had no obvious objections to Mendelism as a theory of heredity he was concerned with its lack of predictive power beyond the offspring of a particular set of parents. Although Mendelism was capable of predicting with certainty the possible types of children of given parents, biometry was more vague but capable of wider application. The probable measurement of

particular organs of the offspring could be calculated from those of the parents and those of the general population, but large numbers of families of similar parents in that population were required before the prediction was accurate. Mendelism dealt with natural pairs of unit characters and biometry with measurements on samples to which one then applied statistical methods. To that extent one might consider biometrical methods as artificially measuring features of a population while Mendelism dealt with natural traits.

Fisher criticized the Mendelians for their preoccupation with detail and an accompanying lack of attention to abstract reasoning — "[a] blindness to well-established, and as far as their own facts are concerned, coordinating principles" [1915, 59]. He claimed that it is in the "highest degree unlikely" that Mendelism will ever cover even the field of heredity, and any amount of information about the mechanism of heredity cannot do away with the need for broad co-ordinating principles of evolution [*Ibid.*, 60]. By contrast, Darwin's work did "cover" the field insofar as it continued to explain and coordinate new facts (the required scientific test). It was in the context of this 1915 work that Fisher introduced an idea that would have an important influence on his thinking about populations and their evolutionary development. The comparison was between populations construed in terms of of Mendelian factors (genes) and the populations of molecules that constitute a gas. The agencies of selection always act amidst a multitude of random causes, each of which may have a predominant influence if we fix our attention on a particular individual. Yet, these agencies determine the progress or decline of the population as a whole. In the case of the kinetic theory we have molecules moving freely in all directions with varying velocities, yet we can obtain a statistical result that is a perfectly definite measurable pressure. Knowledge of the nature and properties of the atom is inessential and independent of our knowledge of general principles in the way that our ability to predict and control the way populations evolve is independent of particular knowledge of individuals. This analogical model would loom large in Fisher's later work, informing the way in which biological populations should be conceived of in order to produce the appropriate kinds of statistical results.

The first step toward a reconciliation of biometry and Mendelism was in 1918 in a paper whose stated goal was the investigation of biometrical properties of a population of a "more general type" than was usually considered, in the hopes that a more exact analysis of the causes of human variability might be given. Fisher began by introducing the elementary statistical concepts involved in measuring variability within a population. Basically if we take a population of roughly 1000 individuals and measure stature, the measurements are usually grouped symmetrically around a mean value, which is the average stature of the sample. The distribution of deviations from this average follows the law of errors allowing us to measure variability as one would measure errors of observation, by the mean of the squares of the deviations of different individuals from the mean. And, as Fisher notes, this mean square deviation is comparable to the mean square error used in the physical sciences. This mean square deviation Fisher called the variance,

where two independent causes of variability acting together produce a variance which is the sum of the variances produced by either separately. Hence, each constituent cause is expressed as a fraction or percentage of the total variance that they jointly produce.

What Fisher wanted to do was determine the extent to which characteristics such as stature were determined by a large number of Mendelian factors. Studies had shown that in the case of brothers the correlation coefficient was around .54, (amount of variance due to ancestry) which leaves 46 percent of the variance to be accounted for in some other way. Both Galton and Pearson had shown that this could not be due to environmental effects and moreover, Galton himself determined that the variance is far less in identical twins. According to the Mendelian hypothesis the large variance among children of the same parents is due to the segregation of those factors in respect to which the parents are heterozygotes. Yule had earlier argued that the effects of dominance and the environment in reducing correlations between relatives were identical, something that Fisher hoped to disprove by separating how much of the total variance was due to dominance, how much resulted from other environmental causes and how much from additive genetic effects. If one could resolve observed variance into these different fractions (i.e. expressing these fractions as functions of observed correlations) then one could easily determine the extent to which nature dominated over nurture. The paper succeeded in showing that the hypothesis of cumulative (multiple) Mendelian factors (genes) fit the observed data and so provided a plausible hypothesis for the inheritance of continuous variates like stature. He demonstrated that the effect of dominance in individual effects expressed itself in a single dominance ratio. And, using fraternal correlation Fisher was able to determine the dominance ratio and distinguish dominance from all non-genetic causes such as environment (which might possibly lower correlations). Essentially what Fisher succeeded in doing was distinguishing not only between genetic and environmental variance but also between the different components of genetic variance itself.

But what exactly was it that enabled Fisher to perform this kind of statistical analysis? Largely it was due to a significant departure from the kind of methodology employed by Pearson and the biometricians. Fisher made a number of explicit assumptions that were clearly at odds with Pearson's earlier work [1904] and [1909]. His main complaint was that Pearson's method was too restrictive with respect to the nature of Mendelian factors. Contra Pearson he did not assume that different Mendelian factors were of equal importance and allowed that different phases of each could occur in any proportions consistent with the conditions of mating. The heterozygote could take any value between dominant or recessive (and even outside that range) resulting in the terms loosing their polarity and becoming simply the means of distinguishing one pure phase from the other. So, all dominant genes did not have a like effect. In order to simplify his calculations Fisher also assumed random mating as well as the independence of the different factors. Finally, and perhaps most importantly, he assumed that the factors were sufficiently numerous so that some small quantities could be neglected.

The simplifying *methodological* assumptions involving independence and an indefinitely large number of Mendelian units were based on the analogy with gas theory that Fisher alluded to in his earlier [1915] work. Essentially he treated large numbers of genes in a way similar to the treatment of large numbers of molecules and atoms in statistical mechanics. By making these simplifying and idealising assumptions Fisher was able to calculate statistical averages that applied to populations of genes in a way *analogous* to calculating the behaviour of molecules that constitute a gas. But, it is important to stress that the analogy was a methodological one. If he could construe populations of genes on analogy with the way statistical mechanics construes populations of molecules, he could perhaps arrive at solutions that might be applicable to the kinds of populations studied by the biometricians — large ones that displayed certain phenotypic characteristics.

Pearson, who acted as a referee for the paper, objected that the hypothesis of an indefinitely large number of Mendelian characteristics was out of the region of experiment using Mendelian methods. Although he was entirely comfortable dealing with large populations (a corner stone of biometric methods), the idea that the manifestation of a character was the result of an *indefinitely* large number of Mendelian factors was anathema. And, while he acknowledged that one could increase the number of factors under consideration, from two to perhaps even four, the idea that these could increase indefinitely was something he took to be amenable to neither proper statistical/biometrical analysis nor experimental tests of the kind favoured by the Mendelians. The other referee, Punnett (a Mendelian), also focused on this assumption. He pointed out that although there were cases worked out for three factors, the mathematics proved very laborious, making it unlikely that Fisher's assumption could ever be tested by experiment. In fact he put the point rather strongly in a remark comparing this kind of work with problems that deal with "weightless elephants on frictionless surfaces, where at the same time we are largely ignorant of the other properties of the said elephants and surfaces" [Norton and Pearson, 1976, 155].

But for Fisher this degree of idealisation was essential to guarantee his method, and hence the legitimacy of its conclusions. In other words, he could escape the difficulties associated with detailed Mendelian analyses by focussing on general principles. But in order to do this it was necessary that he assume a large number of factors in order to establish statistically the generality and validity of the principles. In the way that one can have knowledge about the properties of gases without detailed knowledge of the molecules and atoms that make up the gas, one could have knowledge of how a population would evolve without knowing the details of the heredity of all individual characteristics. And, an indefinite number of traits was essential to the process of averaging that yielded such knowledge.

The important issue here concerns the correct way to model populations. The notion of a population as a way of thinking about groups of organisms had become, for the biometricians, a cornerstone of Darwinian evolutionary theory. Pearson had specific views about how populations ought to be characterised if one was to be able to test the extent to which Mendelism accorded with biometrical findings.

There were a variety of assumptions required for the specification of each of the individual Mendelian factors:

- which allelomorph was dominant
- to what extent did dominance occur
- what were the relative magnitudes of the effects produced by different factors
- in what proportion did the allelomorphs occur in the general population
- were the factors dimorphic or polymorphic, to what extent were they coupled etc.

If one assumed, as Fisher did, an indefinite number of Mendelian factors then the nature of the population could not be specified in any complete sense; thereby undermining any statistical result that might follow. In addition to these assumptions there were the more general considerations that needed to be taken into account regarding homogomy (preferential mating) as opposed to random mating, selection and environmental effects, all of which needed to be treated separately if one was to determine the genetic basis of the inheritance of particular characteristics. So, not only did Fisher differ from Pearson with respect to specific assumptions about the *nature* of Mendelian factors (that all were equally important, etc.) but the way in which one characterised or 'modelled' a Mendelian population was also much more general.

On the view advocated by the biometricians one needs to have variability at the level of individuals as a basis for blending inheritance; but in order to predict how populations will evolve one begins by averaging over this variability to arrive at a statistical characterization of the population. By contrast, Fisher's model of a Mendelian population was based on the molecular models of statistical mechanics. The idealised nature of the assumptions from that domain served as a methodological model or analogy on which he based his own views about how to characterise a Mendelian population. Knowledge of individuals simply wasn't important; hence the idea that each was a source of variation had no role to play in the general conception of the population; variation came at the level of Mendelian characters. But this kind of idealisation or abstraction over the domain of individuals could not be accommodated in Pearson's model world. His objections were not simply to the presence of idealising assumptions since his own statistical methods and the process of averaging obviously involved idealisations as well. Instead, what Pearson objected to was the *kinds* of idealisations Fisher used. Detailed knowledge of the Mendelian factors that characterized the individuals in the population was not considered. Instead one could simply *assume* an indefinite number of such factors and then average over these. But, as we saw above, Pearson thought that many different specifications were required for each factor, and because these factors were understood as the defining characteristics of the individual it was anathema to Pearson that they not be explicitly specified.

4 GENES AND SELECTION: THE SYNTHESIS ESTABLISHED

Although Fisher's synthesis of Mendelian genetics with biometrical Darwinism marked a turning point in the debate with Pearson disagreements of a somewhat different kind began to emerge with the biologist Sewall Wright, another of the founders of population genetics. Wright [1921] also showed how selection operated in Mendelian populations and did so not only by using a different quantitative method but also by incorporating an entirely different qualitative theory about the way selection operated. Wright was convinced by his experimental work with guinea pigs that it was the interaction systems of genes, rather than single genes, that were the important elements of evolution. He also believed, contrary to Fisher, that natural selection operated most effectively in smaller populations where inbreeding was sufficiently intense to create new interaction systems as a result of random drift, but not intense enough to cause random non-adaptive fixation of genes. In those populations natural selection could act on the newly produced interaction systems, resulting in a more rapidly changing population than that produced by mass selection of single genes (as described by Fisher's approach). Wright felt that there was sufficient evidence from animal breeders that mass selection was slow and somewhat unsure; hence he assumed that natural populations tended to become subdivided into partially isolated groups small enough to cause random drifting of genes. Thus, one of the key differences between Fisher and Wright in attempting to isolate the way selection worked concerned population structure. Fisher was convinced that it worked more efficiently in large groups in which there was mass selection and random breeding, whereas Wright's account favoured small inbred groups in which obvious variation would then serve as the basis for selection.

Wright's evolutionary models and mathematical techniques were intended to provide a way of focusing on a quantitative analysis of inbreeding. His use of the method of path coefficients (a system he devised) enabled him to determine the degree to which a given effect was produced by each of a number of different causes. In that analysis the causes of factors that affected a trait (e.g.the general size of an organism) and factors that affected its separate parts could be treated as independent. A path coefficient could then be defined as the ratio of the variability of the effect that is found when all causes are kept constant except the one in question to the total variability. The path coefficient is identical with a partial-regression coefficient provided that it can be measured in standard-deviation units. Wright, however, wanted to stress the difference between dealing with multiple regressions and dealing with his complex network of relations that involved hypothetical factors. As far as he was concerned, correlation and regression coefficients provided descriptive/predictive statistics from which scientists unjustifiably inferred the existence of causal relationships. By contrast, the method of path analysis was not used to *infer* a causal scheme or *deduce* particular causal relationships; instead, it required a specific causal hypothesis to be stated at the outset and was best applied in cases where causal pathways were known or could reasonably be assumed,

but their relative importance was unknown.

Wright was able to extend his method to include cases where causes were correlated instead of independent. He used it to establish the relative importance of heredity and environment in determining piebald patterns in guinea pigs, and also applied it to systems of mating. By looking at the effects of various systems of inbreeding on the genetic composition of the population, the effects of assortive mating, and the effects of selection, Wright was able to calculate the percentages of homozygosity in successive generations. Like those of Fisher, Wright's mathematical methods allowed for precise quantitative determination of the effects of natural selection as the agent of evolutionary change.

Around the same time Fisher [1922] presented an extended discussion of the conditions under which variance could be maintained and a way to determine the overall effects of selection and mutation in achieving this. He wanted to show how gene frequencies would change under selection pressures and particular environmental conditions. Necessary for that approach was the reduction of a population to their composite genes and an isolation of selection pressures from other types of conditions. One again the analogy with gas theory was a crucial piece of the methodological puzzle. By using the model of an ideal gas, Fisher was able to "create" a population in which he could measure effects not measurable experimentally. The mathematical technique used for characterizing a population in the 1918 paper provided the instrument for investigating the role of selection in human populations by replacing actual populations with idealized ones. In the 1922 paper "On the Dominance Ratio", Fisher argued that an equation representing the stochastic distribution of Mendelian determinants in a population over time was the key to an accurate and quantitative understanding of evolution in that population. He needed only general statistical laws about the behaviour among individuals, rather than specific knowledge of the individuals themselves, in order to determine the effects of evolutionary mechanisms.

The paper began with a discussion of equilibrium under selection. Fisher first demonstrated that the frequency ratio for the alleles of a Mendelian factor was a stable equilibrium only if selection favoured the heterozygotes. He then showed that the survival of an individual mutant gene depended on chance rather than selection. Only when large numbers of individuals were affected would the effect of selection override random survival, and even then only a small minority of the population would be affected. Fisher also examined the distribution of factors not acted on by selection, cases of gene extinction counterbalanced by mutation, and extinction in the absence of mutation and selection, where one saw a steady decline in variation due to the effects of random survival (Hagedoorn effect). On the basis of his calculations of the number of genes exterminated in any one generation and the distribution of factors in successive generations he was able to show that even in a population of roughly 10,000 random-breeding individuals without new mutations the rate of gene extinction was extremely small. Hence, the chance elimination of genes could not be considered more important than elimination by selection.

The other important component in the analogy with gas theory was the fact that the distribution of the frequency ratio for different Mendelian factors was calculable from the fact that the distribution was stable in the absence of selection, random survival effects and so forth. Again, the source was the velocity distribution law in gas theory. Just as the formulation of this law assumed an independence of molecules in the gas, so too, Fisher assumed the independence of various hereditary factors from each other and their independence from the effects of selection and random survival. The important difference was that he specified a population in which he could first calculate the distributions without selection, mutation and random survival, and then he used those results to go on and determine how the effects of selection operated in different contexts. To do that Fisher considered the cases of uniform genetic selection in the absence of dominance and genotypic selection with complete dominance. From those distributions he was able to calculate the amount of mutation needed to maintain the variability given a specific amount of selection. To maintain variability in the case of equilibrium in the absence of selection, the rate of mutation had to be increased by a very large quantity. So the presence of even the slightest amount of selection in large populations had considerably more influence in keeping variability in check than did random survival. Consequently, the assumption of genotypic selection balanced by occasional mutations fit the facts deduced from the correlations of relatives in humans.

So, by making simplifying assumptions about the large size of the population and its high degree of genetic variability, Fisher was able to demonstrate how his stochastic distributions led to the conclusion that natural selection acting on single genes (rather than mutation, random extinction, epistasis etc.) was the primary determinant in the evolutionary process. He found that mutation rates significantly higher than any observed in nature could be balanced by very small selection rates. The distribution of the frequency ratios for different factors was calculated from the assumption that the distribution was stable. The kind of statistical independence that figured prominently in the velocity-distribution law was applicable to the effects of selection in Mendelian populations.

As more variables were added, the mathematics Fisher used became intractable, a situation that led to the development of his famous "fundamental theorem of natural selection", published in 1930 in *The Genetical Theory of Natural Selection*. The theorem states that the rate of increase in fitness for any organism at any time is equal to its genetic variance in fitness at that time. The meaning of the theorem and what it implies has been the subject of great controversy in the literature since its publication. It is now widely thought that the correct interpretation of the theorem has been given by Price [1972] and Ewens [1989]. I will not go into the details of the technicalities here but instead will simply give an overview of the theorem so that the reader can have some appreciation of its importance. Perhaps the first thing that needs to be mentioned is that although the theorem refers to the fitness of an organism Fisher intended this to refer to the mean fitness of the species conceived of as an idealised panmictic population. Secondly, the genetic

variance in fitness is in fact the variation resulting from the additive effects of genes. But the very notion of fitness itself was also reconceptualized by Fisher. In his presentation he designated a number of quantities as representing particular aspects of a population. m is the Malthusian parameter referring to the growth rate of the population, M is the mean of this parameter and measures the fitness of the population; W is the genetic variation in fitness, also described as the rate of increase in fitness due to natural selection. C is a constant representing the relation between fitness and population increase and D specifies the rate of loss due to deterioration of the environment. Fisher described the relation among these quantities as follows:

$$dM/dt + M/C = W\text{-}D.$$

While this doesn't actually state the fundamental theorem, the relationship is important for understanding what the theorem actually says. Fisher explicitly claims that the rate of increase of overall fitness at any given moment dM/dt is the result of a balance between the additive genetic variance in fitness W and the totality of effects contributed by the environment. His goal was to isolate within an evolving population *the* particular feature of it that could be identified solely with the operation of natural selection. In order to do this he made use of the methodology developed in 1918 and 1922 that allowed him to analyse the variance of a character into the additive effects of genes, dominance, epistasis and environmental effects. What the theorem states then is an equality between something that is constantly being maximized (the fitness due to selection only) and a variance.

As Price [1972] has pointed out, it has long been a mystery what exactly Fisher meant by his fundamental theorem and how he derived it. Adding to the problem is the fact that most of the discussion and statement of the theorem is done in non-mathematical language; hence one has to rely on the precision of the definitions he uses. As Price points out, part of the confusion results from two different accounts of the theorem. One refers to the rate of increase in fitness due to all changes in gene ratio and the other to the rate of increase in the average value of the Malthusian parameter ascribable to natural selection. In the former case we must include the effects of dominance, epistasis etc. while in the latter we have only the process of selection which involves the distribution of gene frequencies alone (the additive effects of genes) and not the interaction between genes. In Fisher's eyes the theorem identified that which was maximised by natural selection, the genetic variance in fitness. The statistical distribution of genes was for him the fundamental language in which all evolutionary processes should be expressed and understood. To summarise then, the rate of increase in fitness due to changes in the gene ratio was equal to the genetic variance of the fitness W that the population exhibited. If fitness was a measure of the ability of a gene to survive and reproduce, then natural selection acted to increase the fitness of the population.

Once again Fisher appealed to a physical model, namely, gas theory, as a way of grounding the regularity with which W could be expected to vary. He claimed that

such regularity was guaranteed by the same circumstances that made a statistical assemblage of particles like a gas bubble obey the gas law (without appreciable deviation) [Fisher, 1958, 34–40]. He went on to compare his fundamental theorem to the second law of thermodynamics: Both are properties of populations that are true regardless of the nature of their individual units, and both are statistical laws, each of which requires continual increases in a measurable quantity (in one case entropy, and in the other fitness). And, he adds, "it is more than a little instructive that so similar a law should hold the supreme position among the biological sciences" [Fisher, 1958, 39]. The key for understanding Fisher's approach is, once again, the use of idealizing assumptions similar to those used in gas theory. The power of his model derived from elimination of parameters such as migration, isolation, genetic recombination and gene interaction. He was able to separate, in a way that could not be done in empirical studies, the key features in populational variation: genetic factors, environmental factors and sampling errors. Fisher's quantitative expression for genetic variance enabled him to show the amount of variation any particular genetic trait could exhibit in a population distinct from environmentally caused fluctuations.

Building on Hardy's work, Fisher pointed out that in Mendelian inheritance there was no inherent tendency for variability to diminish over time. Alternative genes were conserved unless their proportions were changed by selection, chance or mutation; and as Fisher's calculations showed, selection was the factor most effective in changing gene frequencies. Mutation could no longer be thought of as the force of evolutionary change, for even very small selective effects could overpower it. We know that the Malthusian parameter m, which represents fitness, has a natural-selection component that must always equal W, the genetic variance in fitness. So for Fisher, m did not simply measure changes in population numbers, but rather the change in the total reproductive value of population members; in other words, it measured the rate of population growth in terms of total reproductive value. Because Fisher was able to give a quantitative expression for W, it was possible to determine the effects of selection as a parameter in the overall determination of m.

Wright wrote a review of the Genetical Theory of Natural Selection in 1930 which set out in very brief form some of his disagreements with Fisher, particularly the characterization of selection. He went on to develop his views and opposition to Fisher in a series of papers written over the next ten years. In conclusion I would like to highlight the nature of these differences, many of which remain the subject of debate today.

5 DEBATING THE DETAILS

It was in the 1930 review and his famous paper of 1931 "Evolution in a Mendelian Population" that Wright presented his ideas about the shifting balance theory of evolution and the notion of an adaptive landscape. If one can imagine a field of visible joint frequencies of all genes spread out in a multidimensional space,

with another dimension measuring degree of fitness, then we can think of the field as comprising several peaks or humps relative to the latter. These humps are the result of factors that influence, either positively or negatively, fitness. For Wright the problem of adaptive evolution concerned the mechanism whereby species could find their way from lower to higher peaks in the adaptive landscape under the influence of selection. If the maximum peak a species reached was not the absolute maximum the only way to enhance fitness would be through something other than the steady pressure of selection. Wright suggested four factors that would contribute to this progress; the changing environment would continually change the peaks; new mutations would add further dimensions to the field thereby allowing new possibilities for advancement; random genetic drift would produce stochastic jumps from one peak to another as would the division of a large population into smaller ones.

We can begin to see here just how significant the differences were between Fisher and Wright. For Fisher there is only one way to think about selection — by taking account of the additive effects of genes, with its power limited only by the availability of new variation in the form of mutation. Although both of these are analysed statistically Fisher's overall view of selection is deterministic in the sense that only it and mutation are responsible for evolutionary change. Wright, on the other hand, introduced drift (sampling error) and migration as factors that also affect gene frequencies. He claims that there "must be some trial and error mechanism on a grand scale by which the species may explore the region surrounding the small portion of the field which it occupies. To evolve, the species must not be under the strict control of natural selection" [1932, 359]. Wright's emphasis on interactive genetic systems and the fact that selection is altered by random variations points unequivocally to a causal understanding of the evolutionary process; an understanding that was represented by his use of path analysis in evaluating information relevant for evolving populations.

Although Fisher's work on dominance indicated his awareness of the importance of interactive genetic systems, his interest was in the long term where he thought the effects of single genes would prove more important. So, except in cases where we are interested in the probability of the survival of an individual new mutant, stochastic effects are simply not significant; natural selection is considered the sole agency in genetic evolution. He emphasised the main effects of single loci rather than complexes of loci and the steady and essentially deterministic increase in the frequency of each allele having a selective advantage. This is not to suggest that selection is deterministic in the traditional sense but only that the action of natural selection is not affected by the presence of random factors. However, if we have small populations of the kind that Wright favoured then each tends to become more or less homozygous. But, because of migration there is an infusion of genes into different populations we get another level of selection that is interpopulational. In that sense there are two levels of selection depending on the structure of the populations with the result that selection can operate on either individual genes or gene complexes. For Fisher, of course, selection at the level of gene complexes

was simply not an option.

Without invoking Wright's adaptive landscape theory we can characterise the two approaches to evolution as atomistic vs. holistic [Gayon, 1999]. In Fisher's case the gene itself is the primary element in adaptive evolution while for Wright it is gene interaction. It is important to stress that this is not simply an issue about the importance of selection in different contexts, but rather about its actual operation. For Fisher, selection is a straightforward affair. Once the genetic raw material has been furnished by mutation, natural selection is regarded as the sole agency in shaping genetic evolution. Because complexes of genes at various loci tend to be broken up by recombination, a stronger emphasis is placed on genes at single loci than on gene complexes. And, the total additive genetic variance is defined as the sum of the constituent one-locus marginal values [1918, 405]. Moreover, an allele having a net selective advantage is destined for fixation as long as that advantage persists; hence, even a minute selective force can have evolutionary consequences. Contrast this with Wright's view and its emphasis on the entire gene complex. As I mentioned above, selection is, in many senses, a deterministic matter on Fisher's account but for Wright the importance of relatively small population sizes is that such deterministic behaviour does not occur. Both random drift and random changes in selective values can move gene frequencies to higher selective peaks. Migration between these small populations also results in interpopulational selection. So, the unit of selection here is the entire gene complex and not the individual alleles which are often seen as having no absolute selective advantage. Most characters were thought to be affected by genes at many loci with most genes influencing several characters.

Although Fisher and Wright were indeed asking the same kinds of questions regarding the role of selection in producing adaptation, each had different answers. Fisher's view was that selection operated best in large randomly breeding populations. Wright, on the other hand, argued that there would be a rapid differentiation of local strains which was in itself non-adaptive, but permitted selective increase or decrease of the numbers in different strains. This would result in a relatively rapid adaptive advance of the species as a whole. For Wright the problem with large populations is that they were subject to slow and often reversible changes with selection while small ones were too likely to decline or become extinct as a result of random fixation of disadvantageous genes. In an intermediate population or one with small subgroups there tends to be partly adaptive and partly non-adaptive radiation among the subgroups causing the successful genes to spread and the others to decline. This, he thought, produced changes in the mean gene frequency of the species as a whole with a much faster rate of evolution [1931, 151]. Both random drift and inbreeding were significant in this model because the latter allowed for complex interactive gene systems to be held together and the former allowed the population to come into contact with genes of greater selective advantage. In a small inbred population where only selection operates the result would be stagnation once a certain level of adaptation was reached.

What the synthesis of Darwinian selection and Mendelian genetics produced

was, among other things, complex mathematical frameworks capable of showing that selection operated in Mendelian populations. Fisher introduced new mathematical methods (stochastic models) to model the change in population gene frequencies as a random process evolving in time. He treated the survival of individual genes by means of a branching process analysed by functional interation and then set up a chain-binomial model and analysed it using a diffusion approximation involving partial differential equations [Edwards, 1994]. Branching processes were used to calculate the survival probability of a new mutant. Sewall Wright introduced the method of path analysis as a way of determining various causal factors involved in the evolution of populations. While debates about the merits of Fisher's vs. Wright's account of the evolutionary process continue today, disagreements regarding those details in no way undermines the significance of their achievement and the role that population genetics and its methods have come to play in understanding fundamental aspects of the natural world.

BIBLIOGRAPHY

[Bateson, 1894] W. Bateson. *Materials for the Study of Variation*. London: Macmillan, 1894.

[Bateson, 1897] W. Bateson. Progress in the study of variation, I. *Science Progress*, 1, 1897.

[Bateson, 1901] W. Bateson. Heredity, differentiation and other conceptions of biology: a consideration of Professor Karl Pearson's paper "On the Principle of Homotyposis". *Proceedings of the Royal Society*, 69: 193–205, 1901.

[Castle, 1903a] W. E. Castle. Mendel's laws of heredity. *Proceedings of the American Academy of Arts and Sciences*, 38: 535–48, 1903a.

[Castle, 1903b] W. E. Castle. The laws of heredity of Galton and Mendel, and some laws governing race and improvement by selection. *Proceedings of the American Academy of Arts and Sciences*, 39: 223–43, 1903b.

[Darwin, 1859] C. Darwin. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, London: Murray, 1859. Facsimile, Cambridge: MA: Harvard Press, 1964.

[Edwards, 1994] A. W. F. Edwards. The fundamental theorem of natural selection. *Biological Review*, 69: 443–474, 1994.

[Ewens, 2004] W. Ewens. *Mathematical Population Genetics, 1. Theoretical Introduction*. New York: Springer, 2004.

[Fisher, 1911] R. A. Fisher. Heredity, comparing the methods of biometry and Mendelism. Paper read to Cambridge University Eugenics Society. Reprinted in Norton and Pearson [1976], pages 155–62, 1911.

[Fisher, 1918] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52: 399–433, 1918.

[Fisher, 1922] R. A. Fisher. On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, 42: 321–41, 1922.

[Fisher, 1924] R. A. Fisher. The biometrical study of heredity. *Eugenics Review*, 16: 189–210, 1924.

[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press. 2^{nd} edition [1958], New York: Dover, 1930.

[Fisher and Stock, 1915] R. A. Fisher and C. S. Stock. Cuenot on pre-adaptation. A criticism. *Eugenics Review*, 7: 46–61, 1915.

[Galton, 1865] F. Galton. Hereditary talent and character. *Macmillan's Magazine*, 12: 157–66 and 318–27, 1865.

[Galton, 1869] F. Galton. *Hereditary Genius*. London, Macmillan, 1869.

[Galton, 1885] F. Galton. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute*, 15: 246–63, 1885.

[Galton, 1889] F. Galton. *Natural Inheritance*, London: Macmillan, 1889.

[Galton, 1892] F. Galton. *Finger Prints*. London: Macmillan, 1892.
[Gayon, 1998] J. Gayon. *Darwinism's Struggle for Survival*. Cambridge: Cambridge University Press, 1998.
[Mendel, 1865] G. Mendel. Versuche uber Pflanzen-Hybriden. *Verhandlungen des Naturforschenden Vereines in Brunn*, 44: 3–47, 1865. English translation: Experiments on plant hybrids. In C. Stern and E. R. Sherwood (eds.), *The Origin of Genetics: A Mendel Source Book*, pages 1–48, San Francisco, Freeman, 1966.
[Norton and Pearson, 1976] B. J. Norton and E. S. Pearson. A note on the background to and refereeing of, R.A. Fisher's 1918 Paper 'On the correlation of relatives on the supposition of Mendelian inheritance'. *Notes and Records of the Royal Society*, 31: 151–62, 1976.
[Pearson, 1895] K. Pearson. Contributions to the mathematical theory of evolution. II. Skew variation in homogenous material. *Philosophical Transactions of the Royal Society of London A*, 186: 343–414, 1895.
[Pearson, 1898] K. Pearson. Mathematical contributions to the theory of evolution. On the law of ancestral heredity. *Proceedings of the Royal Society*, 62: 386–412, 1898.
[Pearson, 1900] K. Pearson. *The Grammar of Science*. Second edition. London: Black, 1900.
[Pearson, 1902] K. Pearson. On the fundamental conceptions of biology. *Biometrica*, 1: 320–44, 1902.
[Pearson, 1904] K. Pearson. Mathematical contributions to the theory of evolution. XII. On a generalized theory of alternative inheritance with special reference to Mendel's laws. *Philosophical Transactions of the Royal Society A*, 203: 53–86, 1904.
[Pearson, 1909a] K. Pearson. The theory of ancestral contributions in heredity. *Proceedings of the Royal Socitey B*, 81: 219–24, 1909.
[Pearson, 1909b] K. Pearson. On the ancestral gametic correlations of a Mendelian population mating at random. *Proceedings of the Royal Society B*, 81: 214–29, 1909.
[Pearson, 1911] K. Pearson. *The Grammar of Science*. Third edition. London: Black, 1911.
[Pearson, 1930] K. Pearson. *The life, letters, and labours of Francis Galton*, 3 Vols. Cambridge: Cambridge University Press, 1930.
[Price, 1972] G. R. Price. Fisher's fundamental theorem made clear. *Annals of Human Genetics*, 36: 129–40, 1972.
[Yule, 1902] U. Yule. Mendel's laws and their probable relations to intra-racial heredity. *New Phytologist*, 1: 193–207 and 222–38, 1902.
[Yule, 1907] U. Yule. On the theory of inheritance of quantitative compound characters on the basis of Mendel's laws — A preliminary note. *Report of the 3^{rd} International Congress on Genetics*. London: Spottiswoode, pages 140–42, 1907.
[Wright, 1922] S. Wright. The effects of inbreeding and crossbreeding on guinea pigs. *U.S. Department of Agriculture Bulletin*, 1090: 1–63; 1121: 1–49, 1922.
[Wright, 1929] S. Wright. Evolution in a Mendelian population. *Anatomical Record*, 44: 287, 1929.
[Wright, 1930] S. Wright. Review of the genetical theory of natural selection. *Journal of Heredity*, 21: 349–56, 1930.
[Wright, 1931] S. Wright. Evolution in Mendelian populations. *Genetics*, 16: 97–159, 1931.
[Wright, 1932] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, 1: 352–66, 1932.

This page intentionally left blank

MAXIMISATION PRINCIPLES IN EVOLUTIONARY BIOLOGY

A. W. F. Edwards

1 INTRODUCTION

Maximisation principles in evolutionary biology (or, more properly, extremum principles, so as to include minimisation as well) fall into two classes — those that seek to explain evolutionary change as a process that maximises fitness in some sense, and those that seek to reconstruct evolutionary history by adopting the phylogenetic tree that minimises the total amount of evolutionary change. Both types have their parallels in the physical sciences, from which they originally drew their inspiration. There are important differences, however, and the naive introduction of extremum principles into biology has tended to obscure rather than clarify the underlying processes and estimation procedures, often leading to controversy. To understand the difficulties it is necessary first to note the use of extremum principles in physics.

2 EXTREMUM PRINCIPLES IN PHYSICS

Extremum principles exert a peculiar fascination in science. The classic physics example is from optics, where Fermat in the 17th century established that when a ray of light passes from a point in one medium through a plane interface to a point in another medium with a different refractive index its route is the one which minimises the time it takes for the light to travel. That is, on various assumptions about the speed of light in the different media, he showed that this solution corresponded to Snel's existing Sine Law of refraction. As a matter of fact modern books on optics find it necessary to frame Fermat's principle rather differently, not using the notion of a minimum at all, but the point for us is that in its original form it solves at least some optical path problems by minimisation. The mathematics works out just right, but that is not to agree with Fermat that 'nature is economical'.

In the 18th and 19th centuries mathematical physics unearthed a whole range of extremum principles of great practical value because they provided a unifying point of view which gave results that are mathematically identical to the older approaches. Ideas like potential energy and its minimisation led to the notions

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

of potentials and fields of force in gravitational and electromagnetic theory. The great 18th century developments in the calculus not only provided the necessary mathematical tools but also encouraged thinking in terms of maxima and minima. No-one supposed any longer that nature was economical. Rather, there was an implicit appeal to the classical concept of parsimony, that the simplest mathematical approach to any problem should be preferred, and this often involved an extremum principle. Ockham's razor cut away the unnecessary mathematics but offered no explanation of nature.

A quite different kind of extremum principle grew up more-or-less simultaneously in statistics in these centuries, the most famous example being the method of least squares originating in the work of Legendre and Gauss. Far from being a way of solving problems with mathematical economy, as is Fermat's principle, it is a method for selecting among probability models that one which most closely fits some observed data (in Gauss's case, the observations of cometary positions). There is an element of arbitrariness in the choice of the criterion to be minimised (here the sum of the squares of the residual errors unexplained by the model) but subsequent research in statistical inference has produced plenty of justifications, the most dominant of which is Fisher's method of maximum likelihood.

Both these types of extremum principle have their analogues in evolutionary studies. First, we shall look at attempts to model evolution as a process that maximises 'fitness'.

3 EVOLUTION AS FITNESS-MAXIMISATION. (1) FISHER'S 'FUNDAMENTAL THEOREM'

An English proverb reminds us that 'Nothing succeeds like success' but it was not until 1930 that anyone put figures to it. In *The Genetical Theory of Natural Selection* published that year R. A. Fisher introduced his 'Fundamental Theorem of Natural Selection'. We shall discuss it in more detail in due course, but for the moment we note that its basis is a simple 'growth-rate' theorem: 'The rate of increase in the growth-rate of a subdivided population is proportional to the variance in growth-rates'. It is obvious that if the subdivisions of a population are growing at different rates then those subdivisions with the fastest rates of growth will increasingly dominate so that the overall growth-rate will increase. ('Growth' could here be negative too, but we describe the simplest case.) In full:

> If a population consists of independent groups each with its own growth-rate (the factor by which it grows in a given time interval), the change in the population growth-rate in this interval is equal to the variance in growth-rates amongst the groups, divided by the population growth-rate. [Edwards, 1994, 2002a]

To prove this, let a population be subdivided into k parts in proportions p_i ($i = 1,2, \ldots, k$) and let the ith part change in size by a factor w_i in unit time

(its 'growth-rate'). At the end of the unit time interval the new proportion of the ith part will be $p_i' = p_i w_i / w$, where $w = \Sigma p_i w_i$ is the overall growth rate. The new overall growth rate $w' = \Sigma p_i' w_i$ is therefore equal to $\Sigma p_i w_i^2 / w$ and the change in the overall growth-rate, $w' - w$, is $(\Sigma p_i w_i^2 - w^2)/w$. But the numerator of this expression is simply the variance of the w_i by definition (say v), being their expected squared value minus their squared expected value. Thus $w' - w = v/w$ and the theorem is proved. The constant of proportionality, $1/w$, is of course just a scaling factor: dividing both sides by w leads to $w'/w - 1 = v/w^2$, showing that the proportionate change is equal to the square of the coefficient of variation. A continuous version of the theorem is easily obtained by proceeding to a limit, the growth in each sub-population then being exponential.

It is to Adam Smith's *An Inquiry into the Nature and Causes of the Wealth of Nations* [1776] that we must turn for the germ of the idea behind the Fundamental Theorem. In Book IV: *Of Systems of Political Economy* he wrote:

> As every individual, therefore, endeavours as much as he can both to employ his capital in the support of domestic industry, and so to direct that industry that its produce may be of the greatest value; every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.

As Sober [1984] remarked, 'The Scottish economists offered a non-biological model in which a selection process improves a population as an unintended consequence of individual optimization', and he saw this as one of the influences in the formation of Darwin's views on the effects of natural selection. We can only speculate on whether Fisher was familiar with these ideas in the 1920s; if so, a possible source would have been the economist J. Maynard Keynes, whose *A Treatise on Probability* was published in 1921 and whom Fisher had known since his undergraduate days [Box, 1978]. He will, however, certainly have read Darwin's comment: 'The larger and more dominant groups thus tend to go on increasing in size; and they consequently supplant many smaller and feebler groups' [Darwin, 1859, 428].

The theorem exactly captures the vague notion that the more variable are the growth-rates the more quickly will the most rapidly-growing parts of the population (or sectors of the economy) come to dominate the rest. Eventually dominance by the fastest will be complete, all variability in growth-rates will have vanished,

and no further increase in the overall growth-rate can occur. It is as though the process of change, or evolution, consumes the variability.

At the time when Fisher entered on his researches the importance of heritable variation as the raw material on which natural selection acts was not well understood, and the prevailing intellectual atmosphere was one of doubt as to whether Darwinian natural selection could account fully for evolutionary change. Furthermore, noone had yet successfully quantified this variation mathematically. Indeed, not until Francis Galton had statisticians learnt to regard biological variability as intrinsically interesting, in contrast to variability in physics and astronomy which they had been accustomed to think of as 'error'. Karl Pearson, in his many papers on evolution, had pursued the mathematical study of biological variation, but it was the young Fisher who saw the overwhelming advantages of working not with the standard deviation but its square, which he christened the *variance* in his paper *The correlation between relatives on the supposition of Mendelian inheritance* [Fisher, 1918a]. As he explained in a contemporary article in the *Eugenics Review*: 'This mean square deviation I term the variance, and use it as a measure of variability, by reason of this important property, namely, that two independent causes of variability acting together produce a variance which is the sum of the variances produced by either separately' [Fisher, 1918b].

Throughout the 1920s, during the gestation of *The Genetical Theory of Natural Selection*, Fisher was fond of quoting Darwin's dictum, from Chapter II of *The Origin of Species*, that 'wide ranging, much diffused, and common species vary most'. 'The present note', he wrote in his major 1922 paper [Fisher, 1922a], 'is designed to discuss the distribution of the frequency ratio of the allelomorphs of dimorphic factors, and the conditions under which the variance of the population may be maintained'. Fisher's investigations were thus motivated by a desire to investigate the importance of the various factors affecting the variance of a population. Stochastic losses due to finite population size will cause a slow decay, but the major contributions will be the gain due to mutation and the loss due to selection: 'The decay in the variance of a species breeding at random without selection, and without mutation, is almost inconceivably slow: a moderate supply of fresh mutations will be sufficient to maintain the variability. When selection is at work even to the most trifling extent, the new mutations must be much more numerous to maintain equilibrium. ... Thus a numerous species, with the same frequency of mutation, will maintain a higher variability than will a less numerous species: in connection with this fact we cannot fail to remember the dictum of Charles Darwin [quoted above]'.

Armed with his appropriate measure of variability and inspired by Darwin's work, it was thus natural that Fisher should proceed to quantify the relationship between the variability in the genetic contributions of individuals to subsequent generations, and the rate at which the population changes in consequence. In 1926, in a note in *Nature* with E. B. Ford, he wrote 'it is easily demonstrable that in species in which a higher proportion of the total variance is ascribable to genetic causes, the effective selection will be more intense than in species in which

the variance is to a larger extent ascribable to environmental variations' [Fisher and Ford, 1926] (repeated in [Fisher and Ford, 1928]).

Calling the relative genetic contributions of individuals their fitnesses, in *The Genetical Theory of Natural Selection* Fisher sought a relationship between the variance in fitness in the population and the fitness of the whole population (as measured by the mean of the individual fitnesses). What he found (in modern terminology; *cf.* [Edwards, 1994]) was that

> The rate of increase in the mean fitness of a population ascribable to natural selection acting through changes in gene frequencies is equal to its additive genetic variance in fitness.

In *The correlation between relatives* Fisher had already shown how the total genotypic variance at a single locus in a population, that is, the variance contributed by the variability of genotypes in an environment assumed uniform, could be considered as the sum of two components. The first component, now called the additive genetic variance, is due to the additive effects of the genes ('an additive part which reflects the genetic nature without distortion'), whilst the second, residual component, the dominance variance, is due to the non-additive effects of the genes. The basic idea comes from linear regression: regress the genotypic values on the number (0, 1 or 2) of genes of a particular allelic type and the resultant analysis of variance separates the linear ('additive genetic') and the residual ('dominance') effects. (For more information consult, for example, the books by Falconer [1989], or Edwards, [2000].)

The Fundamental Theorem can be viewed as an adaptation of the growth-rate theorem to the particular needs of population genetics (though it is in fact anachronistic to do so). It tells us how the changes in gene frequencies brought about by natural selection affect the mean fitness.

4 EVOLUTION AS FITNESS-MAXIMISATION. (2) MISINTERPRETING THE FUNDAMENTAL THEOREM

It may seem, and has seemed to many, that a theorem which gives the rate at which the mean fitness of a population increases must imply that the course of evolution can be charted by elucidating the route that would increase the fitness as rapidly as possible, a triumphant mathematization of the 'invisible hand'. In the genetical context this is not the case, for two reasons.

In the first place, this is not the theorem Fisher proved. The unravelling of the Fundamental Theorem took many years (see [Edwards, 1994; 2002b], for the history) but it is now entirely clear that it refers only to the change in that component of the overall fitness that can be ascribed directly to gene-frequency change. Moreover, Fisher embedded his theorem in a discussion in *The Genetical Theory* that emphasizes that this change is only one part of the change in fitness, and that other factors such as mutation and changes in the environment are continuously at

work to influence the overall mean fitness of a population, which must always be roughly constant for a population to remain at about the same size. It is as though Fisher's theorem has been interpreted as a kind of 'big-bang' theory of evolution whereas it is in fact more descriptive of a 'steady-state' evolutionary universe.

Secondly, even taking the naive interpretation of the theorem as exactly true for the simplest case of a single genetic locus with additive fitnesses, it still does not follow that gene-frequency change will be such as to maximise the increase in fitness. The mathematics simply does not imply this as a consequence. (We look at what can in fact be said in the next section.)

Unfortunately, two powerful strains of misunderstanding corrupted the interpretation of the theorem. The first is quickly disposed of. It is easy to find examples of populations modelled mathematically in which it is not true that the overall mean fitness increases; ironically, the first was put forward by Fisher himself [1941] in a failed attempt to rescue his theorem from misinterpretation. In consequence many commentators, mainly American, have wrongly maintained that the theorem had been disproved by counter-example and could therefore be only approximately true at best.

The second misunderstanding was Sewall Wright's conception of an 'adaptive landscape' of fitness up which populations would move, like hill-walkers ascending a peak by the steepest route. It lent itself to the extension, important in Wright's view of evolution, that this landscape was complex and multi-peaked and that an important element of change was the ability of populations to make stochastic jumps from one place to another, thus enabling valleys to be crossed and higher peaks of fitness attained. Wright eventually admitted that his supposition that progress would be by the steepest route could not be justified mathematically (at least, not in the normal gene-frequency space — see the next section), but the whole development, which became the dominant paradigm in evolutionary genetics, derived originally from his reading of *The Genetical Theory* and the Fundamental Theorem.

Wright had called the mean fitness of a population 'W', and Fisher's last sally against it was scathing:

> I have never indeed written about W and its relationships, and now that the alleged relationship has been brought to my attention, I must point out that the existence of such a "potential function" as that which Wright designates by W, is not a general property of natural populations, but arises only from the special and restricted cases which Wright has chosen to consider. Selective tendencies are not, in general, analogous to what mechanicians describe as a conservative system of forces. To assume this property is one of the gravest faults of Wright's formulation.

(For the source of this quotation, and all historical points, see [Edwards, 1994]; for a critique of Wright's 'shifting balance' theory of evolution and its stochastic component, see [Coyne *et al.*, 1997].)

5 EVOLUTION AS FITNESS-MAXIMISATION. (3) WHAT REMAINS TO BE SAID

If it is true that the course of evolution cannot in general be described in terms of a potential function like those that exist in physics, is it possible to salvage anything from the idea at all in the context of evolution?

First, a distinction must be made between complete 'dynamic' theories involving a potential function that can fully describe the processes of change, and weaker 'static' theories of maximisation which merely describe stationary points of a system in terms of the maximisation of some non-decreasing function. The latter class includes the basic multiallelic single-locus model of population genetics in which stable equilibria are at a maximum of the mean fitness, as proved by a number of authors around 1960 (see [Edwards, 2000a], for the history). However, since the proof is specific to the problem no general theory emerges.

Somewhere between the dynamic and the static theories there is the possibility that changes can be described by a variational principle, but not one that involves a potential function. This turns out to be the case for the above model provided a transformed gene-frequency space is adopted, as follows.

Svirezhev [1972] stated that in the continuous-time model for selection at a multiallelic locus the application of Fisher's angular transformation $\sin^{-1}\sqrt{p}$ to the gene frequencies results in (1) the direction of the gene-frequency change being given by the steepest direction on the function W and (2) the magnitude of the change being proportional to the square root of the additive genetic variance, all with respect to the new coordinate system. This does not in itself define a potential function, though the case of only two alleles exhibits an interesting atypical feature in the discrete-generation case because the square root of the additive genetic variance is exactly proportional to the slope (taken positively) of the mean fitness W in the transformed space [Edwards, 2000b]. Thus the magnitude of the change is indeed proportional to the slope and W is then a potential function.

It is possible to pursue this line of approach using the Riemannian distance $\Sigma[(dp_i)^2/p_i]$ implied by Fisher's transformation, but it requires advanced mathematics and the reader is referred to the books by Svirezhev and Passekov [1990], Hofbauer and Sigmund [1998] and Bürger [2000]. However, in a more accessible account Ewens [1992] placed the result in the context of an optimization principle originally proposed by Kimura [1958] and developed by Ewens to prove that natural selection changes the gene frequencies so as to maximise the partial increase in the mean fitness subject to a constraint on the Riemannian distance between new and old gene frequencies. This is indeed a maximisation principle, but does not go so far as to establish a potential function.

We thus see that a naive description of evolution as a process that tends to increase fitness is misleading in general, and hill-climbing metaphors are too crude to encompass the complexities of Mendelian segregation and other biological phenomena. In some particular cases there is a deep structure involving variational principles, but it is principally of mathematical interest only. Fisher's Fundamen-

tal Theorem of Natural Selection seems to be about as much as it is possible to say about the relationship between genetical variation and evolutionary change, but it does not lead to any grand theory such as is to be found in the physical sciences.

6 EVOLUTION AS ENTROPY MAXIMISATION

Before leaving the topic of theories of evolution involving maximisation it should just be mentioned that occasionally authors become excited at the possibility of viewing evolution in the language of entropy. Fisher himself was not immune to this in *The Genetical Theory*: he likens the Fundamental Theorem to the second law of thermodynamics in some important respects, especially that 'each requires the constant increase of a measurable quantity, in the one case the entropy of a physical system and in the other the fitness ... of a biological population'.

It seems that he was here carried away by his enthusiasm, for a qualification quickly follows. Amongst a list of 'profound differences': '(3) Fitness may be increased or decreased by changes in the environment ...'. Moreover, '(4) Entropy changes are exceptional in the physical world in being irreversible, while irreversible evolutionary changes form no exception among biological phenomena'. 'Finally, (5) entropy changes lead to a progressive disorganization of the physical world ... while evolutionary changes are generally recognized as producing progressively higher organization in the organic world'.

In making the comparison Fisher was here relying on his 'mathematical education [which] lay in the field of mathematical physics', as he once wrote (see [Edwards, 2002c]). He had spent a year in the Cavendish Laboratory, Cambridge, after his first degree.

Subsequent authors have retraced Fisher's footsteps with rather less caution and complete ignorance of the Fundamental Theorem. As an example we may cite the collection of papers entitled *Entropy, Information and Evolution: New perspectives on physical and biological evolution* [Weber *et al.*, 1988] in which some twenty names have managed to write a whole book about *entropy*, about *information*, and about *evolution*, without once discussing Fisher's writings on them.

More recently *Fisher information* has been invoked to develop the mathematics of population genetics [Frieden *et al.*, 2001]. Fisher information is a statistical concept distinct from, and earlier than, *Shannon information* in communication theory. Whereas Shannon information measures the capacity of a channel to transmit a message, Fisher information measures the informativeness of a body of data, or of a statistic derived from data, about the parameter of the underlying probability model. Shannon is concerned with the medium, Fisher with the message. All three concepts are intimately connected with the *likelihood*. Shannon information is minus the expected log-likelihood function at the true value of the parameter and Fisher information is the second derivative of this with respect to the parameter, and measures the sharpness of the peak of the log-likelihood (see, for example, [Edwards, 1992]). The basis of the suggestion that Fisher information might be

used to develop population genetics is that it has the same mathematical form as the Reimannian distance metric mentioned in the last section, but Edwards [2002c] has argued that this is only of superficial interest.

7 RECONSTRUCTING EVOLUTION: THE PRINCIPLE OF PARSIMONY AND OCKHAM'S RAZOR

We now turn away from principles that emulate the techniques of mathematical physics to one that involves a minimisation principle which ultimately takes its justification from the theory of statistics and in particular the method of maximum likelihood. The problem it addresses is that of reconstructing the course of evolution from a knowledge of present-day characters — be they morphological traits amongst species, blood-group frequencies amongst populations or DNA sequences amongst individuals. The heuristic foundation is that similar things are closely related. But why?

In *The Origin of Species* Darwin [1859] replaced the hypothesis of the individual creation of each species with the hypothesis of a single origin of life followed by the evolution of new species. He did not appeal to Ockham's razor in support of the change, and would probably have been surprised to have been told that his achievement was an example — perhaps the example par excellence — of the application of Ockham's razor. His argument, rather, was one of analogy with the way lineages in genealogies could be traced back to common ancestors.

Ockham's razor, sometimes now called *the principle of parsimony*, is traditionally associated with William of Ockham (*ca.* 1280-1349):

> *Pluralitas non est ponenda sine necessitate* (multiplicity ought not to be posited unnecessarily).

In recent years this principle has sometimes been invoked by systematists to justify particular methods of phylogenetic systematics (for reviews see [Sober, 1988], and [Stewart, 1993]). In this context the principle manifests itself as what we might call 'the hypothesis of the single origin', whether of life itself or of a particular character. Darwin drew his first evolutionary tree with a single root in 1837, and the principle that the hypothesis of a single origin is to be preferred where possible can be called *the Darwin principle*. Thus if a particular character appears the same, or very similar, in two species, we are to prefer the hypothesis of a common origin for this character over the hypothesis of two independent origins and parallel selection, unless, of course, other evidence overwhelms it. 'Perhaps the correct way of viewing the whole subject', wrote Darwin in *The Origin*, 'would be, to look at the inheritance of every character whatever as the rule, and non-inheritance as the anomaly.' Again, in 1862 he wrote 'Is it not a more simple and intelligible view that all the Orchideæ owe what they have in common, to descent from some monocotylodenous plant . . . ?' [Darwin, 1862].

Ockham's razor invites us to prefer the hypothesis of the single origin, but the rise of probabilistic thinking in the eighteenth century rendered this explanation

for the preference obsolete. The probabilistic revolution applied the calculus of probability, invented in connection with games of chance, to the problem of the observational errors of astronomy and geodesy. The failure of a hypothesis to account exactly for observational data was no longer seen as a fault of either the hypothesis or the data, but as a measurable reflection of the random component in the observations. *Anceps fortuna aequitate rationis reprimitur* Pascal had written in 1654 — 'Proper calculation masters fickle fortune' (see [Edwards, 2002d]). People still wanted simple hypotheses, but they were now probabilistic ones, and needed to be judged not only by simplicity but by a definite measure of how well they explained the data — by 'likelihood'.

An example will make clear how probabilistic reasoning leads naturally to the Darwin principle. Suppose we hear that a friend has borne male twins. Are they identical? If so, the probability that they are both boys is 1/2, but if they are not identical the probability is 1/4 (assuming that boys occur with probability 1/2). Setting aside for the moment any question of the prior probabilities, the likelihood ratio in favour of the twins being identical is therefore 2:1. The hypothesis of the single origin of the male sex is better supported than the hypothesis of two separate origins. Now suppose that the character about which we are told is not sex but a much rarer genetic character segregating in families with low frequency p. The twins both have this character. Are they identical?

This time the probability that they both possess the character is p if they are identical, but p^2 if they are not. The likelihood ratio in favour of identical twins is therefore $p{:}p^2$ or $1{:}p$, which is large and convincing if p is small. The more improbable the event that has been observed, the greater the support for the hypothesis of a single origin. The word 'likelihood' is here used in its technical statistical sense defined by Fisher [1921]. A full account of it will be found in Edwards [1972; 1992].

The Darwin principle was, of course, in use in connection with human similarities millenia before Darwin. Noone ever suggested that each individual was a special creation like a pre-Darwinian species (souls were another matter), and inferring relationship from similarity of appearance must always have reflected a subconscious desire to avoid having to account independently for similar facial and other features. Darwin himself, in a rare mention of the 'doctrine of chances', was more explicit [1859], unwittingly using a likelihood argument for the complementary problem of inferring inheritance given a known relationship:

> When a deviation appears not unfrequently, and we see it in the father and child, we cannot tell whether it may not be due to the same original cause acting on both; but when amongst individuals, apparently exposed to the same conditions, any very rare deviation, due to some extraordinary combination of circumstances, appears in the parent — say, once amongst several million individuals — and it reappears in the child, the mere doctrine of chances almost compels us to attribute its reappearance to inheritance.

Nowadays human geneticists implicitly use the same argument to infer relationship given the mode of inheritance: if a rare gene is found in two people in a small community the instant reaction is to enquire whether they are related.

8 RECONSTRUCTING EVOLUTION: THE METHOD OF MINIMUM EVOLUTION

In 1963 Cavalli-Sforza and Edwards suggested the following principle for estimating the shape and dimensions of a phylogenetic tree:

THE PRINCIPLE OF MINIMUM EVOLUTION

The most plausible estimate of the evolutionary tree is that which invokes the minimum nett [*sic*] amount of evolution.

They thus made explicit what had for so long been implicit in phylogenetic thinking, and they did so because the existence of the new electronic computers which they were harnessing to the problem compelled the replacement of general ideas like Ockham's razor and the Darwin principle with unambiguous concepts which were *programmable*. As soon as objects that are to be connected by a phylogenetic tree can be displayed in a multi-dimensional character-space (it need not even be Euclidean — with DNA sequences, for example, it is a lattice; [Edwards and Cavalli-Sforza, 1964]) the construction of a tree connecting them with the minimum total edge-length is a programmable problem. The first example of a solution was presented at the International Congress of Genetics in 1963 [Cavalli-Sforza and Edwards, 1964].

Cavalli-Sforza and Edwards were quick to claim that the justification for their 'method of minimum evolution', as it became known, was statistical, based on a probability model for the evolutionary process coupled with standard statistical estimation of the tree. Having originally used the method of least squares, they soon developed a detailed probability model which enabled them to apply the method of maximum likelihood. They then underpinned their method of minimum evolution with the argument that it gave results which were close to the maximum-likelihood solution, thus again reflecting the link between traditional phylogenetic procedures and their justification in terms of likelihood. The history has been given by Edwards [1996], while the book *Inferring Phylogenies* by Felsenstein [2004] is a rich source of information about the many statistical procedures which have been developed in the past forty years.

Unfortunately some of the terminology has changed in the course of time. Camin and Sokal [1965] introduced the phrase 'principle of evolutionary parsimony' but it inevitably became shortened, in the phylogenetic context, to 'principle of parsimony', thus confusing it with Ockham's razor, while 'method of minimum evolution' was used in a new (and now dominant) sense by Rzhetsky and Nei [1993].

The three common procedures currently in use, minimisation of the total tree length ('parsimony'), least squares applied to a matrix of pairwise distances, and

maximum likelihood applied to a specific model for the evolutionary process, are all examples of the use of extremum principles. But whereas the attempt to apply the extremum methods of mathematical physics to the study of evolutionary *processes* has proved largely unsuccessful because the parallels are insufficiently close — a clear case of analogical seduction — the application of statistical extremum methods, especially maximum likelihood, to the study of evolutionary *inferences* has proved exceptionally rewarding.

BIBLIOGRAPHY

[Box, 1978] J. F. Box. *R.A.Fisher: the Life of a Scientist.* New York: Wiley, 1978.

[Bürger, 2000] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation.* Chichester, U.K.: Wiley, 2000.

[Camin and Sokal, 1965] J. H. Camin and R. R. Sokal. A method for deducing branching sequences in phylogeny. *Evolution*, 19: 311–326, 1965.

[Cavalli-Sforza and Edwards, 1964] L. L. Cavalli-Sforza and A. W. F. Edwards. Analysis of human evolution. *Genetics Today: Proc. 11th Int. Congress of Genetics, The Hague, 1963*, 3: 923–933, 1964.

[Coyne *et al.*, 1997] J. A. Coyne, N. H. Barton, and M. Turelli. Perspective: A critique of Sewall Wright's shifting balance theory of evolution. *Evolution*, 51: 643–671, 1997.

[Darwin, 1859] C. Darwin. *On the Origin of Species by means of Natural Selection.* London: Murray, 1859.

[Darwin, 1862] C. Darwin. *The Various Contrivances by which Orchids are Fertilised by Insects.* London: Murray.

[Edwards, 1992] A. W. F. Edwards. *Likelihood.* Baltimore: Johns Hopkins University Press, 1992.

[Edwards, 1994] A. W. F. Edwards. The fundamental theorem of natural selection. *Biological Reviews*, 69: 443–474, 1994.

[Edwards, 1996] A. W. F. Edwards. The origin and early development of the method of minimum evolution for the reconstruction of evolutionary trees. *Systematic Biology*, 45: 79–91, 1996.

[Edwards, 2000a] A. W. F. Edwards. *Foundations of Mathematical Genetics* (2nd ed.). Cambridge University Press, 2000a.

[Edwards, 2000b] A. W. F. Edwards. Sewall Wright's equation $\Delta q = (q(1\text{-}q)\partial w/\partial q)/2w$. *Theoretical Population Biology*, 57: 67–70, 2000b.

[Edwards, 2002a] A. W. F. Edwards. Fundamental Theorem of Natural Selection. *Encyclopedia of Evolution.* New York: Oxford University Press, 371–372, 2002a.

[Edwards, 2002b] A. W. F. Edwards. Fundamental theorem of natural selection. *Theoretical Population Biology*, 61: 335–7, 2002b.

[Edwards, 2002c] A. W. F. Edwards. Fisher information and the fundamental theorem of natural selection. *Rendiconti Scienze B*, 134: 1–15 (2000), *Istituto Lombardo di Scienze e Lettere, Milano*, 2002c.

[Edwards, 2002d] A. W. F. Edwards. *Pascal's Arithmetical Triangle.* Baltimore: Johns Hopkins University Press, 2002d.

[Edwards and Cavalli-Sforza, 1964] A. W. F. Edwards and L. L. Cavalli-Sforza. Reconstruction of evolutionary *trees. Systematics Assoc. Publ. No. 6, Phenetic and phylogenetic classification*, 67–76, 1964.

[Ewens, 1992] W. J. Ewens. An optimizing principle of natural selection in evolutionary population genetics. *Theor. Pop. Biol.*, 42: 333–346, 1992.

[Falconer, 1989] D. S. Falconer. *Introduction to Quantitative Genetics* (3rd ed.). London: Longman, 1989.

[Felsenstein, 2004] J. Felsenstein. *Inferring phylogenies.* Sunderland, Massachusetts: Sinauer, 2004.

[Fisher, 1918a] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52: 399–433, 1918a.

[Fisher, 1918b] R. A. Fisher. The causes of human variability. *Eugenics Review*, 10: 213–20, 1918b.
[Fisher, 1922a] R. A. Fisher. On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, 42: 321–41, 1922a.
[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930; 2nd ed: New York, Dover Publications, 1958; variorum ed: Oxford University Press, 1999.
[Fisher and Ford, 1926] R. A. Fisher and E. B. Ford. Variability of species. *Nature*, 118: 515–516, 1926.
[Fisher and Ford, 1928] R. A. Fisher and E. B. Ford. The variability of species in the Lepidoptera, with reference to abundance and sex. *Trans. Roy. Entomological Soc. Lond.*, 76: 367–379, 1928.
[Frieden *et al.*, 2001] B. R. Frieden, A. Plastino, and B. H. Soffer. Population genetics from an information perspective. *Journal of Theoretical Biology*, 208: 49–64, 2001.
[Hofbauer and Sigmund, 1998] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.
[Keynes, 1921] J. M. Keynes. *A Treatise on Probability*. London: Macmillan, 1921.
[Kimura, 1958] M. Kimura. On the change of population fitness by natural selection. *Heredity*, 12: 145–167, 1958.
[Rzhetsky and Nei, 1993] A. Rzhetsky and M. Nei. Theoretical foundations of the minimum-evolution method of phylogenetic inference. *Molecular Biology and Evolution*, 10: 1073–1095, 1993.
[Smith, 1776] A. Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Strahan and Cadell, 1776.
[Sober, 1984] E. Sober. *The Nature of Selection*. Cambridge, Massachusetts: MIT Press, 1984.
[Sober, 1988] E. Sober. *Reconstructing the past: parsimony, evolution, and inference*. Cambridge, Massachusetts: MIT Press, 1988.
[Stewart, 1993] C-B. Stewart. The powers and pitfalls of parsimony. *Nature*, 361: 603–607, 1993.
[Svirezhev, 1972] Y. M. Svirezhev. [*In Russian*] Optimum principles in population genetics. *Studies on Theoretical Genetics*, Novosibirsk, 86–102, 1972.
[Svirezhev and Passekov, 1990] Y. M. Svirezhev and V. P. Passekov. *Fundamentals of Mathematical Evolutionary Genetics*. Dordrecht: Kluwer, 1990.
[Weber *et al.*, 1988] B. H. Weber, D. J. Depew, and J. D. Smith (eds). *Entropy, Information, and Evolution: New perspectives on physical and biological evolution*. Cambridge, Mass.: MIT Press, 1988.

This page intentionally left blank

REDUCTIONISM IN BIOLOGY

Alex Rosenberg

Physicalism about biology is the thesis that all facts, including all the non-macromolecular biological facts, are fixed by the facts of molecular biology. Reductions argue that physicalism mandates that non-molecular biological explanations need to be completed, corrected, made more precise or otherwise deepened by more fundamental explanations in molecular biology. Antireductionism does not dispute physicalism's metaphysical claim about the fixing of biological facts by macromolecular ones, but denies it has implications either for explanatory strategies or methodological morals. The antireductionists holds that explanations in functional biology need not be corrected, completed or otherwise made more adequate by explanations in terms of molecular biology.

1 POST-POSITIVIST INTERTHEORETICAL REDUCTION

Reduction was supposed by the post-Logical Empiricists to be a relation between theories. In Ernest Nagel's *Structure of Science* [1961], reduction is characterized by the deductive derivation of the laws of the reduced theory from the laws of the reducing theory. The deductive derivation requires that the concepts, categories and explanatory properties, or natural kinds of the reduced theory be captured in the reducing theory. To do so, terms that figure in both theories must share common meanings. Though often stated explicitly, this second requirement is actually redundant as valid deductive derivation presupposes univocality of the language in which the theories are expressed. However, as exponents of reduction noted, the most difficult and creative part of a reduction is establishing these connections of meaning, i.e. formulating "bridge principles", "bi-lateral reduction sentences", "coordinating definitions" that link the concepts of the two theories. Thus it was worth stating the second requirement explicitly. Indeed, early and vigorous opponents of reduction as the pattern of scientific change and theoretical progress argued that the key concepts of successive theories are in fact incommensurable in meaning, as we shall see immediately below.

Within a few years Watson and Crick's Nobel for uncovering the structure and function of DNA, reductionists began to apply their analysis to the putative reduction of Mendelian or population genetics to molecular genetics. The difficulties they encountered in pressing Watson and Crick's discovery into the mold of theoretical reduction became a sort of "poster-child" for antireductionists. In an early and insightful contribution to the discussion of reduction in genetics Kenneth Schaffner observed that reduced theories are usually less accurate and less

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

complete in various ways than reducing theories, and therefore incompatible with them in predictions and explanations. Accordingly, following Schaffner, the requirement was explicitly added that the reduced theory needs to be "corrected" before its derivation from the reducing theory can be effected. This raised a problem which became non-trivial in the fall-out from Thomas Kuhn's *Structure of Scientific Revolutions* [1961], and Paul Feyerabend's "Reduction, Empiricism and Laws" [1964]. It became evident in these works that "correction" sometimes resulted in an entirely new theory, whose derivation from the reducing theory showed nothing about the relation between the original pair. Feyerabend's examples were Aristotelian mechanics, Newtonian mechanics, and Relativistic mechanics, whose respective crucial terms, 'impetus' and 'inertia', 'absolute mass' and 'relativistic mass' could not be connected in the way reduction required. No one has ever succeeded in providing the distinction that reductionism required between 'corrections' and 'replacements'.

More fundamentally, reductionism as a thesis about formal logical relations among theories was undermined by the increasing dissatisfaction among philosophers of science with the powers of mathematical logic to illuminate interesting and important methodological matters such as explanation, and theory testing. Once philosophers of science began to doubt whether deduction from laws was always sufficient or necessary for explanation, the conclusion that inter-theoretical explanation need take the form of reduction was weakened. Similarly, reductionism is closely tied to the axiomatic or so-called syntactic approach to theories, an approach which explicates logical relations among theories by treating them as axiomatic systems expressed in natural or artificial languages. But for a variety of reasons, the syntactic approach to theories has given way among many philosophers of biology to the so-called 'semantic' approach to theories. The semantic approach treats theories not as axiomatic systems in artificial languages, but as sets of closely related mathematical models. On the semantic view the reduction of one theory to another is a matter of employing (one or more) model(s) among those which constitute the more fundamental theory to explain why each of the models in the less fundamental theory are good approximations to some empirical processes, showing where and why they fail to be good approximations in other cases. The models of the more fundamental theory can do this to the degree that they are realized by processes that underlie the phenomena realized by the models of the less fundamental or reduced theory. There is little scope in this sort of reduction for satisfying the criteria for post-Positivist reduction.

To the general philosophical difficulties which the post-positivist account of reduction faced, biology provided further distinct obstacles. David Hull [1973] was the first to notice it is difficult actually to define the term 'gene' as it figures in functional biology by employing only concepts from molecular biology. The required "bridge principles" between the concept of gene as it figures in population biology, evolutionary biology, and elsewhere in functional biology and as it figures in molecular biology could not be constructed. And all the ways philosophers contrived to preserve the truth of the claim that the gene is nothing but a (set of)

string(s) of nucleic acid bases, could not provide the systematic link between the functional 'gene' and the macromolecular 'genes' required by a reduction. There is of course no trouble identifying 'tokens' — particular bits of matter we can point to–of the population biologist's genes with 'tokens' of the molecular biologist's genes. But token-identities won't suffice for reduction, even if they are enough for physicalism to be true. The second problem facing reductionism in biology is the absence of laws, either at the level of the reducing theory or the reduced theory, or between them. If there aren't any laws in either theory, there is no scope for reduction at all.

Whereas the antireductionists were at most able to show that the criterion of connectability with respect to the Mendelian and the molecular gene was not fulfilled as the two theories were in fact stated, we can go much further in vindicating their conclusion. We can demonstrate that the criterion required by "layer-cake" reductionism cannot be satisfied as a fundamental matter of biological process. Individuation of types in biology is almost always via function: to call something a wing, or a fin, or a gene is to identify it in terms of its function. But biological functions are naturally-selected effects. And natural selection for adaptations — i.e. environmentally appropriate effects — is blind to differences in physical structure that have the same or roughly similar effects. Natural selection "chooses" variants by *some of their effects*, those which fortuitously enhance survival and reproduction. When natural selection encourages variants to become packaged together into larger units, the adaptations become functions. Selection for adaptation and function kicks in at a relatively low level in the organization of matter. Accordingly, the structural diversity of the tokens of a given Mendelian or classical or population biological or generally 'functional' gene will mean that there is no single molecular structure or manageably finite number of sets of structures that can be identified with a single functional gene.

Philosophers will recognize the relationship between the functional gene and the DNA sequence as one of "multiple realization" common to the relation functionalism in the philosophy of psychology alleges to obtain between psychological states and neural ones. The blindness of selection for effects to differences in structure provides the explanation for why multiple realization obtains between genes and polynucleotide molecules. Indeed almost every functional kind in biology will be multiply realized, owing to the fact that the kind has an evolutionary etiology.

Functional biology tells us that there is a hemoglobin gene, and yet there is no unique sequence of nucleic acids that is identical to this hemoglobin gene — nothing that could provide a macromolecular definition of the hemoglobin gene of functional biology. Of course there is some ungainly disjunction of all the actual ways nucleic acid sequences nowadays do realize or in the past have realized the hemoglobin gene — i.e. all the sequences that can be translated, and transcribed into RNA which in a local ribosome will code for one or another of the different types — fetal, adult, or the varying defective hemoglobin protein sequences. But this ungainly disjunction, even if we knew it, and we don't, won't serve to define the functional hemoglobin gene. The reason is obvious to the molecular biologist.

An even vaster disjunction of nucleic acid sequences than the actual sequence will work just as well, or indeed just as poorly to constitute the functional hemoglobin gene (and probably will do so in the future, given environmental contingencies and mutational randomness). Just think of the alternative introns that could separate exon regions of the sequence (and may do so in the future, given mutation and variation). And then there are all the promoter and repressor genes, and their alternative sequences, not to mention the genes for producing the relevant ribosomal protein-synthesizing organelles, all equally necessary for the production of the hemoglobin protein, and so claiming as much right to be parts of the functional hemoglobin gene as the primary sequence of the coding region of structural gene itself. Just as the actual disjunction is too complex to state, and yet not biologically exhaustive of the ways to code for a working hemoglobin protein, so also all these other contributory sequences don't exhaust the actual biological alternatives, and so make the macromolecular definition of the functional hemoglobin gene a will-o-the wisp.

In other words, being a hemoglobin molecule "supervenes", in the philosopher's term, on being a particular sequence of amino acids, even though there is no complete specification possible or scientifically fruitful of all the alternative particular sequences of amino acids that could constitute (i.e. realize) the function of the hemoglobin molecule in oxygen transport. Roughly, a biological property, P, supervenes on some (presumably complex) physical and/or chemical property or other, Q, if and only if anything else that has physical/chemical property Q, must also have biological property P, (see [Rosenberg, 1978]). There is among philosophers a fairly sustained debate about the force of the "must" in this formulation. Does the supervenience of the biological on the physical/chemical have to obtain in virtue of natural laws, or even some stronger sort of metaphysical necessity. As many philosophers argue, biological properties are "local" — make implicit but ineliminable reference to a particular place and time (the Earth since 3.5 billion years ago). Thus, it may be that biological properties are only locally supervenient, a much weaker thesis than one which makes it a matter of general law every where and always in the universe, (see "Concepts of Supervenience" in [Kim, 1993]).

When a biological property is supervenient on more than one complex physical/chemical property then it is also a multiply realized property. The supervenience of the biological on the physical is a way of expressing the thesis of physicalism. The blindness of natural selection to differences in structure is what turns the supervenience of the biological on the physical into the multiple realization of the biological by the physical. This structural diversity explains why no simple identification of molecular genes with the genes of population genetics of the sort post-positivist reduction requires is possible. More generally, the reason there are no laws in biology is thus the same reason there are no bridge-principles of the sort post-positivist reduction requires (This result will be even less surprising in light of the post-positivist realization that most bridge principles in science will be laws, not definitions.)

The unavoidable conclusion is that as far as the post-positivist or "layer-cake" model of intertheoretical reduction is concerned, none of its characteristic preconditions are to be found in theories of functional biology, theories of molecular biology, or for that matter in any future correction of one or the other of these theories.

2 INTERTHEORETICAL ANTIREDUCTIONISM

If antireductionism were merely the denial that post-positivist reduction obtains among theories in biology, it would be obviously true. But recall, antireductionism is not merely a negative claim. It is the thesis that

a) there are generalizations at the level of functional biology,

b) these generalizations are explanatory,

c) there are no further generalizations outside of functional biology which explain the generalizations of functional biology,

d) there are no further generalizations outside functional biology which explain better, more completely, or more fully, what the generalizations of functional biology explain.

All four components of antireductionism are daunted by at least some of the same problems that vex reductionism: the lack of laws in functional biology and the problems facing an account of explanation in terms of derivation from laws. If there are no laws and/or explanation is not a matter of subsumption, then antireductionism is false too. But besides the false presuppositions antireductionism may share with reductionism, it has distinct problems of its own. Indeed, these problems stem from the very core of the antireductionist's argument, the appeal to ultimate explanations under written by the theory of natural selection.

To see the distinctive problems that an appeal to the ultimate/proximate distinction raises for biology's autonomy, consider a paradigm of putative irreducible functional explanation advanced by antireductionists. Our example is due to one of the most influential of antireductionist physicalists, Philip Kitcher. It is one that has gone largely unchallenged in the almost two decades between the first and the latest occasion in which it has been invoked in his rejection of reductionism. The example is the biologist's explanation of independent assortment of functional genes. The explanadum is:

(G) Genes on different chromosomes, or sufficiently far apart on the same chromosome, assort independently.

According to Kitcher, the functional biologist proffers an explanans for (G), which we shall call (PS):

(PS) Consider the following kind of process, a *PS*-process (for *pairing* and *separation*). There are some basic entities that come in pairs. For each pair, there is a correspondence relation between the parts of one member of the pair and the parts of the other member. At the first stage of the process, the entities are placed in an *arena.* While they are in the arena, they can exchange segments, so that the parts of one member of a pair are replaced by the corresponding parts of the other members, and conversely. After exactly one round of exchanges, one and only one member of each pair is drawn from the arena and placed in the *winners box.* In any PS-process, the chances that small segments that belong to members of different pairs or that are sufficiently far apart on members of the same pair will be found in the winners box are independent of one another. (G) holds because the distribution of chromosomes to games at meiosis is a PS-process.

Kitcher writes, "This I submit is a full explanation of (G), and explanation that prescinds entirely from the stuff that genes are made of". [Kitcher, 1999, 199–200]

Leave aside for the moment the claim that (PS) is a full explanation of (G), and consider why, according to the antireductionist, no molecular explanation of (PS) is possible. The reason is basically the same story we learned above about why the kinds of functional biology cannot be identified with those of molecular biology. Because the same functional role can be realized by a diversity of structures, and because natural selection encourages this diversity, the full macromolecular explanation for (PS) or for (G) will have to advert to a range of physical systems that realize independent assortment in many different ways. These different ways will be an unmanageable disjunction of alternatives so great that we will not be able to recognize what they have in common, if indeed they do have something in common beyond the fact that each of them will generate (G). Even though we all agree that (G) obtains in virtue only of molecular facts, nevertheless, we can see that because of their number and heterogeneity, these facts will not supplant (PS)'s explanation of (G), or for that matter supplant (G)'s explanation of particular cases of genetic recombination. This is supposed to vindicate antireductionism's theses that functional explanations are complete and that functional generalizations cannot be explained by non-functional ones, nor replaced by them.

But this argument leaves several hostages to fortune. Begin with (G). If the argument of the previous section is right, (G) is not a law at all, but the report of a conjunction of particular facts about a spatiotemporally restricted kind, "chromosomes" of which there are only a finite number extant over a limited time period at one spatio-temporal region (the Earth). Accordingly, (G) is not something which we can expect to be reduced to the laws of a more fundamental theory, and the failure to do so constitutes no argument against reductionism classically conceived, nor is the absence or impossibility of such a reduction much of an argument **for** antireductionism.

The antireductionist may counter that regardless of whether (G) is a nomic generalization, it has explanatory power and therefore is a fit test-case for reduction.

This, however, raises the real problem which daunts antireductionism. Antireductionism requires an account of explanation to vindicate its claims. Biologists certainly do accord explanatory power to (G). But how does (G) explain? And the same questions are raised by the other components of the antireductionist's claims. Thus: What certifies (PS) — the account of PS-processes given above — as explanatory? What prevents the vast disjunction of macromolecular accounts of the underlying mechanism of meiosis from explaining (PS) or for that matter from explaining (G) and indeed whatever it is that (G) explains?

There is one tempting answer to this question due to Putnam [1975, 295–298] and Garfinkel [1981], that is widely popular among antireductionists. This is the square peg-round hole argument. On this view, explanations of why a particular square peg goes through the round hole in a board based on considerations from geometry are superior to explanation of the same event that advert to quantum mechanics; the former explanations are entirely adequate and correct, and require no supplementation, correction, or deepening by more fundamental considerations about the material composition of the peg and board, or laws and generalizations that they instantiate.

The reason given for this conclusion is that the latter explanation provides irrelevant detail and fails to identify features of the expanandum that are shared with other similar cases. This argument makes the adequacy of one explanation and the inadequacy of the other turn on whether information is "relevant", not relevant only to the causal process involved, but relevant to something else as well — presumably our interests. No one could deny that the material composition of peg and board, and the laws of governing it are causally relevant to the explanandum. If relevance is to be judged by other criteria, These other criteria will include our interests. Similarly, the similarity of the explanandum to others must be understood as relative. The appeal to similarity also drags in our interests, as Wittgenstein first noticed.

Reductionists are inclined to argue that explanations of why square pegs don't go through round holes which advert to geometry only are either seriously incomplete or false: We need to add information that assures us of the rigidity of the materials under the conditions that obtain when the peg is pushed through the hole, and once we begin trying to make our explanation complete and correct, the relevance of the more fundamental physical facts and laws governing them becomes clearer. Sober [1999] advances a slightly different argument against Putnam's conclusion that the geometrical explanation is superior, which however has a conclusion similar to the Reductionist's. He notes that Putnam's argument begins by conceding that both explanations are correct, or at least equally well supported. Accordingly he infers that the only reason Putnam can offer for preferring the broader, geometrical explanation to the deeper physical one is our "subjective" interests. Putnam would be better advised simply to deny that the quantum theoretical description of the causal process instantiated by the peg and whole is explanatory at all. But it is hard to see how one could disqualify the quantum story as not explanatory at all, even if it were guilty of irrelevant de-

tail and silence about an objective pattern instantiated by this and other peg-and whole cases.

Kitcher implicitly exploits the Putnam/Garfinkel approach in answering the question of why a macromolecular explanation of (PS) is not on the cards? One answer is presumably that it is beyond the cognitive powers of any human contemplating the vast disjunction of differing macromolecular processes each of which gives rise to meiosis, to recognize that conjoined they constitute an explanation of (PS). Or similarly, it is beyond the competence of biologists to recognize how each of these macromolecular processes gives rise to (G). That the disjunction of this set of macromolecular processes implements PS-processes and thus bring about (PS) and (G) does not seem to be at issue. Only some one who denied the thesis of physicalism — that the physical facts fix all the biological facts — could deny the causal relevance of this vast motley of disparate macromolecular processes to the existence of (PS) and the truth of (G).

In fact, there is something that the vast disjunction of macromolecular realizations of (PS) have in common that would enable the conjunction of them to fully explain (PS) to someone with a good enough memory for details. Each was selected for because each implements a PS process and PS processes are adaptive in the local environment of the Earth from about the onset of the sexually reproducing species to their extinction. Since selection for implementing PS processes is blind to differences in macromolecular structures with the same or similar effects, there may turn out to be nothing else completely common and peculiar to all macromolecular implementations of meiosis besides their being selected for implementing PS processes. But this will be a reason to hold that each of these macromolecular implementations explains PS and/or G, and on something other than a pragmatic, erotetic or interest-relative theory of explanation

Antireductionists who adopt what is called an erotetic account of explanation, in preference to a unification account, a causal account or the traditional D-N account of explanation, will feel the attractions of the Putnam/Garfinkel approach. For the erotetic account of explanations treats them as answers to "why questions" posed about a particular occurrence or state of affairs, which are adequate — i.e. explanatory — to the degree they are appropriate to the back-ground information of those who pose the why-question and to the degree that the putative explanation excludes competing occurrences or states of affairs from obtaining. Since it may be that we never know enough for a macromolecular answer to the question of why does (G) obtain, no macromolecular explanation of why (G) obtains will be possible. Similarly, we may never know enough for a macromolecular explanation of (PS) to be an answer to our question "Why do PS processes occur?" But this seems a hollow victory for antireductionism, even if we grant the tendentious claim that we will never know enough for such explanations to succeed. What is worse, it relegates antireductionism to the status of a claim about biologists, not about biology. Such philosophical limitations on our epistemic powers have been repeatedly breeched in the history of science.

Antireductionists wedded to alternative, non-erotetic accounts of explanation, cannot adopt the gambit of a Putnam/Garfinkel theory of explanation in any case. They will need a different argument for the claim that neither (G) nor (PS) can be explained by its macromolecular supervenience base (see pages 366–67 above), and for the claim that (PS) does explain (G) and (G) does explain individual cases of recombination. One argument such Antireductionists might offer for the former claim rests on a metaphysical thesis: that there are no disjunctive properties or that if there are, such properties have no causal powers. Here is how the argument might proceed: The vast motley of alternative macromolecular mechanisms that realize (PS) have nothing in common. There is no property–and in particular no property with the causal power to bring about the truth of (G) which they have in common. Physicalism (which all Antireductionists party to this debate embrace) assures us that whenever PS obtains, some physical process, call it P_i, obtains. Thus we can construct the identity (or at least the bi-conditional) that

$$(R)\mathrm{PS} = P_{1,} \vee P_2 \vee \ldots \vee P_{i,} \vee \ldots \vee P_m,$$

where m is the number, a very large number, of all the ways macromolecular processes can realize PS processes.

The Putnam/Garfinkel theory of explanation tells us that (R) is not explanatory roughly because it's too long a sentence for people to keep in their heads. A causal theory of explanation might rule out R as explaining PS on the ground that the disjunction, $P_1, \vee P_2 \vee \ldots \vee P_{i,} \vee \ldots \vee P_m,$, is not the **full** cause. This might be either because it was incomplete — there is always the possibility of still another macromolecular realization of PS arising, or because disjunctive properties just aren't causes, have no causal powers, perhaps aren't really properties at all. A unificationist-theory of explanation (or for that matter a D-N account) might hold that since the disjunction cannot be completed, it will not effect deductive unifications or systematizations. Thus (PS) and (G) are the best and most complete explanations biology can aspire to. Antireductionist versions of all three theories, the causal, the unificationist, and the erotetic/pragmatic need the disjunction in (R) to remain uncompleted in order to head off an reductionist explanation of (PS) and/or (G).

Consider the first alternative, that (R) is not complete, either because some disjuncts haven't occurred yet or perhaps that there are an indefinite number of possible macromolecular implementations for (PS). This in fact seems to me to be true, just in virtue of the fact that natural selection is continually searching the space of alternative adaptations and counter-adaptations, and that threats to the integrity and effectiveness of meiosis might in the future result in new macromolecular implementations of (PS) being selected for. But this no concession to antireductionism. It is part of an argument that neither (PS) nor (G) report an explanatory generalization, that they are in fact temporarily true claims about local conditions on the Earth.

On the second alternative, (R) can be completed in principle, perhaps because there are only a finite number of ways of realizing a (PS) process. But the dis-

junction is not a causal or a real property at all. Therefore it cannot figure in an explanation of either (PS) or (G). There are several problems with such an argument. First, the disjuncts in the disjunction of $P_1, \vee P_2 \vee \ldots \vee P_i, \vee \ldots \vee P_m$, do seem to have at least one or perhaps even two relevant properties in common: each was selected for implementing (PS) and causally brings about the truth of (G). Second, we need to distinguish predicates in languages from properties in objects. It might well be that in the language employed to express biological theory, the only predicate we employ that is true of every P_i is a disjunctive one, but it does not follow that the property picked out by the disjunctive predicate is a disjunctive property. Philosophy long ago learned to distinguish things from the terms we hit upon to describe them.

How might one argue against the causal efficacy of disjunctive properties? One might hold that disjunctive properties will be causally efficacious only when their disjuncts subsume similar sorts of possible causal processes. If we adopt this principle, the question at issue becomes one of whether the disjunction of $P_1, \vee P_2 \vee \ldots \vee P_i, \vee \ldots \vee P_m$ subsumes similar sorts of causal processes/ The answer to this question seems to be that the disjunction shares in common the features of having been selected for resulting in the same outcome — PS processes. Thus, the disjunctive predicate names a causal property, a natural kind. Antireductionists are hard pressed to deny the truth and the explanatory power of(R).

Besides its problems in undermining putative macromolecular explanations of (PS), (G) and what (G) explains, antireductionism faces some problems in substantiating its claims that (PS) explains (G) and (G) explains individual cases of genetic recombination. The problems of course stem from the fact that neither (PS) nor (G) are laws, and therefore an account is owing of how statements like these can explain.

3 HISTORICAL REDUCTIONISM

Both the "layer-cake" reductionism of post-positivist philosophers of science and its antireductionist rejection are irrelevant to the real issue about the relation between non-molecular biology and molecular biology. If there is a real dispute here, it is not about the derivability or underivability of laws in functional biology from laws in molecular biology, as there are no laws in either subsdicipline. Nor can the real dispute turn on the relationship between theories in molecular and functional biology. There is only one general theory in biology, Darwinism. As Dobzhansky recognized, it is equally indispensable to functional and molecular biology. Once this conclusion is clear, the question of what was reductionism in the post positivist past can be replaced by the question of what reductionism is now. For the obsolescence of the posit-positivist model of reduction hardly makes the question of reductionism or its denial obsolete. The accelerating pace of developments in molecular biology makes this question is more pressing than ever. But it is now clear that the question has to be reformulated if it is to make contact with real issues in biology.

Biology is unavoidably terrestrial. Its explanatory resources are spatiotemporally restricted in their meanings. Thus, the debate between reductionism and antireductionism will have to be one about the explanation of particular historical facts, some obtaining for longer than others, but all of them ultimately the contingent results of general laws of natural selection operating on boundary conditions. Reductionism needs to claim that the most complete, correct, and adequate explanations of historical facts uncovered in functional biology is by appeal to other historical facts uncovered in molecular biology, plus some laws that operate at the level of molecular biology. Antireductionism must claim that there are at least some explanations in functional biology which cannot be completed, corrected, or otherwise improved by adducing wholly nonfunctional considerations from molecular biology. One way to do this would be show that there are some functional biological phenomena that cannot in principle be decomposed or analyzed into component molecular processes. But such a demonstration would threaten the antireductionist's commitment to physicalism. A more powerful argument for antireductionism would be one which shows that even in macromolecular explanations, there is an unavoidable commitment to ultimate explanation by (implicit) appeal irreducible functional — i.e. evolutionary laws, such as the PNS.

Reductionists can provide a strong argument for their view and rebut antireductionist counterargument effectively. But to do so they need to show that ultimate explanations in functional biology are unavoidably inadequate, and inadequate in ways that can only be improved by proximate explanations from molecular biology. This would indeed refute antireductionism. Or it would do so if the reductionist can show that these proximate explanations are not just disguised ultimate explanations themselves. It is the literal truth of Dobzhanky's dictum that of course will make one suspicious that this cannot be done. What the reductionist must ultimately argue is that the laws of natural selection to which even their most macromolecular explanations implicitly advert, are reducible to laws of physical science. This second challenge is the gravest one reductionism faces. For if at the basement level of molecular biology there is to be found a general law not reducible to laws of physics and chemistry, then antireductionism will be vindicated at the very core of the reductionist's favored subdiscipline.

Let us consider the first challenge, that of showing what makes ultimate explanations in functional biology inadequate in ways only proximate molecular explanations can correct. Recall Mayr's [1981] distinction between proximate and ultimate explanation. Consider the question of why butterflies have eye-spots on their wings. This question may express a request for an adaptationist explanation that accords a function, in camouflage for instance, to the eye-spot on butterfly wings, or it may be the request for an explanation of why at a certain point in development eye-spots appear on individual butterfly wings and remain there throughout their individual lives. Reductionism in biology turns out to be the radical thesis that ultimate explanations must give way to proximate ones and that these latter will be molecular explanations.

To expound this thesis about explanations, reductionism adduces another dis-

tinction among explanations.. The distinction is between what are called "how-possibly explanations" and "why-necessary explanations." How-possible explanations show how something could have happened, by adducing facts which show that there is after all no good reason for supposing it could not have happened. A why-necessary explanation shows that its explanandum had to have happened. These two different kinds are distinct and independent of one another. Each kind of explanation will be appropriate to a different inquiry, even when the two different inquiries are expressed in the same words.

There is an important asymmetry between how-possible and why-necessary explanations that philosophers of history recognized. Once a how–possible explanation has been given, it makes perfect sense to go on and ask for a why-necessary explanation. But the reverse is not the case. Once a why-necessary explanation has been given, there is no point asking for a how-possible explanation. For in showing why something had to happen, we have removed all obstacles to its possibly happening. Some philosophers of history went on to suggest that why necessary explanation are "complete". But this is a notion hard to make clear in the case of, say, causal explanations, in which it is impossible to describe all the conditions, positive and negative, individually necessary and jointly sufficient for the occurrence of an event which we seek to explain. For our purposes all that will be required is the observation that a why-necessary explanation provides more information about exactly how its explanans came about than a how-possible explanation, and that is the source of the asymmetry between them. It is not difficult to graft this distinction on to the one broached above between erotetic and pragmatic approaches to explanation. On the erotetic view, whether a question expresses a request for a how-possible explanation or a why-necessary one is a matter of the context in which the question is put, the information available to the interlocutors, their aims and interests. Accordingly, sometimes a why-necessary explanation will not be an appropriate response to an explanatory question. But all this is compatible with the fact that a why necessary explanation provides more information about the causally necessary conditions for the matter to be explained. The exponent of a non-erotetic approach to explanations will hold that there is such a thing as a complete and correct explanation independent of contexts of inquirer's questions, and that insofar as they are both incomplete, the how-possible explanation is more incomplete and the why-necessary closer to the whole story. The reductionist will sympathize with this view, as we shall now see.

Consider the ultimate explanation for eyespots in the buckeye butterfly species *Precis coenia*. Notice to begin with there is no scope for explaining the law that these butterflies have eye-spots, or patterns that may include eye spots, scalloped color patterns, or edge-bands, even though almost all of them do have such markings. There is no such law to be explained, as there are no laws about butterflies, still less any species of them. That the buckeye butterfly has such eyespots is however a historical fact to be explained.

The ultimate explanation has it that eyespots on butterfly and moth wings have been selected for over a long course of evolutionary history. On some butterflies

these spots attract the attention and focus the attacks of predators onto parts of the butterfly less vulnerable to injury. Such spots are more likely to be torn off than more vulnerable parts of the body, and this loss does the moth or butterfly little damage, while allowing it to escape. On other butterflies, and especially moths, wings and eye spots have also been selected for taking the appearance of an owl's head, brows and eyes. Since the owl is a predator of those birds which consume butterflies and moths, this adaptation provides particularly effective camouflage.

Here past events help to explain current events via implicit principles of natural selection. Such ultimate explanations have been famously criticized as "just-so" stories, allegedly too easy to frame and too difficult to test [Gould and Lewontin, 1979] and though its importance has been exaggerated, there is certainly something to this charge. Just because available data or even experience shows that eyespots are wide-spread does not guarantee that they are adaptive now. Even if they are adaptive now, this is by itself insufficient grounds to claim they were selected because they were the best available adaptation for camouflage, as opposed to some other function or for that matter that they were not selected at all but are mere "spandrels", or traits riding piggy-back on some other means of predator avoidance or some other adaptive trait.

Reductionists will reply to this criticism that adaptationalist ultimate explanations of functional traits are "how-possibly" explanations, and the "just-so-story" charge laid against ultimate explanation on these grounds mistakes incompleteness (and perhaps fallibility) for untestability. The reductionist has no difficulty with the ultimate functional how-possibly explanation, as far as it goes. For its methodological role is partly one of showing how high fitness *could* in principle be the result of purely non-purposive processes. More importantly, on the reductionist's view, such a how-possibly explanation sets the research agenda which seeks to provide why-necessary explanations. It is these why-necessary explanations which cash in the promissory notes offered by the how-possibly explanation. But if we are not already convinced reductionists we may well ask, why must such why-necessary explanations be macromolecular? The reason is to be found in a limitation on ultimate explanations recognized by many: its silence about crucial links in the causal chains to which it adverts.

The how-possibly explanation leaves unexplained several biologically pressing issues, ones which are implicit in biologically well-informed request for an ultimate explanation. These are the question of what alternative adaptive strategies were available to various lineages of organisms, and which were not, and the further question of how the feed-back from adaptedness of functional traits–like the eyespot–to their greater subsequent representation in descendants was actually effected. The most disturbing lacuna in a how-possibly explanation is its silence on the causal details of exactly which feed-back loops operate from fortuitous adaptedness of traits in one or more distantly past generations to improved adaptation in later generations and how such feed-back loops approach the biological fact to be explained as a locally constrained optimal design. Dissatisfaction with such explanations, as voiced by those suspicious of the theory of natural selection, those

amazed by the degree of apparent optimality of natural design, as well as the religious creationist, all stem from a single widely shared and very reasonable scientific commitment. It is the commitment to complete causal chains, along with the denial of action at a distance, and the denial of backward causation. Long before Darwin, or Paley for that matter, Spinoza diagnosed the defect of purposive or goal directed explanation: it "reverses the order of nature", and makes the cause the effect. Natural selection replaces goal-directed processes. But natural selection at the functional level is silent on the crucial links in the causal chain which convert the appearance of goal-directedness into the reality of efficient causation. Therefore, explanations that appeal to it sometimes appear to be purposive or give hostages to fortune, by leaving too many links in their causal chains unspecified. Darwin's search for a theory of heredity reflected his own recognition of this fact.

The charge that adaptational explanations are unfalsifiable or otherwise scientifically deficient reflects the persistent claim by advocates of the adequacy of ultimate explanations that their silence on these details is not problematic. A macromolecular account of the process can answer these questions. Such an account would itself also be an adaptational explanation: it would identify strategies available for adaptation by identify the genes (or other macromolecular replicators) which determine the characteristics of Lepidopterans evolutionary ancestors, and which provide the only stock of phenotypes on which selection can operate to move along pathways to alternative predation-avoiding outcomes — leaf color camouflage, spot-camouflage, or other forms of Batesian mimicry, repellant taste to predators, Mullerian mimicry of bad tasting species, etc. The reductionist's "why-necessary explanation" would show how the extended phenotypes of these genes competed and how the genes which generated the eyespot eventually become predominant, i.e. are selected for. In other words, the reductionist holds that a) every functional ultimate explanation is a how-possibly explanation, and b) there is a genic and biochemical path-way selection process underlying the functional how-possibly explanation. As we shall see below, reduction turns the merely how-possible scenario of the functional ultimate explanation in to a why-necessary proximate explanation of a historical pattern. Note that the reductionist's full explanation is still a historical explanation in which further historical facts — about genes and pathways — are added, and are connected together by the same principles of natural selection that are invoked by the ultimate functional how-possibly explanation. But the links in the causal chain of natural selection are filled in to show how past adaptations were available for and shaped into today's functions.

Antireductionist will differ from reductionists not on the facts but on whether the initial explanation was merely an incomplete one or just a how-possibly explanation. Antireductionists will agree that the macromolecular genetic and biochemical path way are causally necessary to the truth of the purely functional ultimate explanation. But they don't complete an otherwise incomplete explanation. They are merely further facets of the situation that molecular research might illuminate. [Kitcher, 1999, 199]. The original ultimate answer to the question why do butterflies have eyespots does provide a complete explanatory answer to a ques-

tion. Accordingly, how-possibly explanations are perfectly acceptable ones, or else the ultimate explanation in question is something more than a mere how-possibly explanation.

Who is right here?

4 COMPLETING WHY-NECESSARY EXPLANATIONS IN EVOLUTIONARY BIOLOGY

On an erotetic view, how-possibly and why-necessarily explanations may be accepted as reflecting differing questions expressed by the same words. The reductionist may admit that there are contexts of inquiry in which how-possible answers to questions satisfy explanatory needs. But the reductionist will insist that in the context of advanced biological inquiry, as opposed say to secondary school biology instruction, for example, the how-possible question either does not arise, or having arose in a past stage of inquiry, no longer does. How-possibly questions do not arise where the phenomena to be explained are not adaptations at all, for instance constraints, or spandrels, and the only assurance that in fact how-possibly explanations make true claims is provided by a why-necessary explanation that cashes in their promissory notes by establishing the adaptive origins of the functional traits in molecular genetics. This will become clearer as we examine proximate explanation in biology.

Consider the proximate explanation from the developmental biology of butterfly wings and their eyespots. Suppose we observe the development of a particular butterfly wing, or for that matter suppose we observe the development of the wing in all the butterflies of the buckeye species, *Precis coenia*. Almost all will show the same sequence of stages beginning with a wing imaginal disk eventuating in a wing with such spots, and a few will show a sequence eventuating in an abnormal wing or one without the characteristic eyespot maladapted to the butterfly's environment. Rarely one may show a novel wing or markings fortuitously more well adapted to the environment than the wings of the vast majority of members of its species.

Let's consider only the first case. We notice in one buckeye caterpillar (or all but a handful) that during development an eyespot appears on the otherwise unmarked and uniform epithelium of the emerging butterfly wing. If we seek an explanation of the sequence in one butterfly, the general statement that in all members of its species development results in the emergence of an eyespot on this part of the wing, is unhelpful. First because examining enough butterflies in the species shows it is false. Second, even with an implicit ceteris paribus clause, or a probabilistic qualification, we know the "generalization" simply describes a distributed historical fact about some organisms on this planet around the present time and for several million years in both directions. One historical fact cannot by itself explain another, especially not if its existence *entails* the existence of the fact to be explained. That all normal wings develop eyespots does not go very far in explaining why one does.

Most non-molecular generalizations in developmental biology are of this kind.

That is, they may summarize sequences of events in the lives of organisms of a species or for that matter in organisms of higher taxa than species. Here is an example of typical generalizations in developmental biology from Wolpert [1997, 320]:

> Both leg and wing discs [in *Drosophila*] are divided by a compartmental boundary that separates them into anterior and posterior developmental region. In the wing disc, a second compartment boundary between the dorsal and ventral regions develops during the second larval instar. When the wings form at metamorphosis, the future ventral surface folds under the dorsal surface in the distal region to form the double layered insect wing.

Despite its singular tone, this is a general claim about all (normal) drosophila embryos, and their leg- and wing-imaginal discs. And it is a purely descriptive account of events in a temporal process recurring in all (normal) *Drosophila* larva. For purposes of proximate explanation of why a double layer of cells is formed in any one particular embryo's imaginal disc, this statement is no help. It simply notes that this happens in them all, or that it does so 'in order' to eventually form the wing, where the 'in order to' is implicit in the small word 'to'.

How is the pattern of eyespot development described in the extract from Wolpert in fact to be proximally explained? Here some of the details of a developmental explanation may be given in order to show its special relevance to the proximate/ultimate distinction. Having identified a series of genes which control wing development in *Drosophila*, biologists then discovered homologies between these genes and genes expressed in butterfly development, and that whereas in the fruit fly they control wing formation, in the butterfly they also control pigmentation. The details are complex but following out a few of them shows us something important about how proximate why-necessary explanation can cash in the promissory notes of how-possibly explanation and in principle reduce ultimate explanations to proximate ones.

In the fruit fly, the wing imaginal disk is first formed as a result of the expression of the gene *wingless* (so called because its deletion results in no wing imaginal disk and no wing) which acts a position signal to cells directing specialization into the wing disc-structure. Subsequently, the homeotic selector gene *apterous* is switched on and produces apterous protein only in the dorsal compartment of the imaginal disk control formation of the dorsal (top) side of the wing and activates two genes, *fringe* and *serrate* which form the wing margin or edge. These effects were discovered by preventing dorsal expression of *apertous,* which results in the appearance of ventral (bottom) cells on the dorsal wing, with a margin between them and other (nonectopic) dorsal cells. Still another gene, *distal-less*, establishes the fruit fly's wing tip. Its expression in the center of the (flat) wing imaginal disk specifies the proximo-distal (closer to body/further from body) axis of wing development. It is the order in which certain genes are expressed, and the concentration of certain proteins in the ovum which explains the appearance of

eye-spot development in the buckeye butterfly.

Once these details were elucidated in *Drosophila* it became possible to determine the expression of homologous genes in other species, in particular in *Precis coenia*. To begin with, nucleic acid sequencing showed that genes with substantially the same sequences were to be found in both species. In the butterfly these homologous genes were shown to also organize and regulated the development of the wing, though in some different ways. For instance, in the fruit fly *wingless* organizes the pattern of wing margins between dorsal and ventral surfaces, restricts the expression of *apterous* to dorsal surfaces and partly controls the proximo-distal access where *distal-less* is expressed. In the butterfly, *wingless* is expressed in all the peripheral cells in the imaginal disk which will not become parts of the wing, where it programs their death. [Nijhout, 1994, p. 45] *Apterous* controls the development of ventral wing surfaces in both fruit flies and butterflies, but the cells in which it is expressed in the *Drosophila* imaginal disk are opposite those in which the gene is expressed in *Precis* imaginal disks. As Nijhout describes the experimental results:

> The most interesting patterns of expression are those of *Distal-less*. In *Drosophila Distal-less* marks the embryonic premordium of imaginal disks and is also expressed in the portions of the larval disk that will form the most apical [wing-tip] structures...In *Precis* larval disks, *Distal-less* marks the center of a presumptive eyespot in the wing color pattern. The cells at this center act as inducers or organizers for development of the eyespot: if these cells are killed, no eyespot develops. If they are excised, and transplanted elsewhere on the wing, they induce an eyespot to develop at an ectopic location around the site of implantation... the pattern of *Distal-less* expression in *Precis* disks changes dramatically in the course of the last larval instar [stage of development]. It begins as broad wedge shaped patters centered between wing veins. These wedges gradually narrow to lines, and a small circular pattern of expression develops at the apex of each line...
>
> What remains to be explained is why only a single circle of *Distal-less* expression eventually stabilizes on the larval wing disks. [Nijhout, p. 45]

In effect, the research program in developmental molecular biology is to identify genes expressed in development, and then to undertake experiments — particularly ectopic gene-expression experiments–which explain the long established observational "regularities" reported in traditional developmental biology. The *explanantia* uncovered are always "singular" boundary conditions insofar as the explananda are spatiotemporally limited patterns, to which there are always exceptions of many different kinds. The reductionistic program in developmental molecular biology is to first explain the wider patterns, and then explain the exceptions — "defects of development" (if they are not already understood from

the various ectopic and gene deletion experiments employed to formulate the why-necessary explanation for the major pattern).

The developmental molecular biologists, S.B. Carroll and colleagues, who reported the beginnings of the proximal explanation sketched above, eventually turned their attention to elucidating the ultimate explanation. Carroll *et al.*, write:

> The eyespots on butterfly wings are a recently derived evolutionary novelty that arose in a subset of the Lepidoptera and play an important role in predator avoidance. The production of the eyespot pattern is controlled by a developmental organizer called the focus, which induces the surrounding cells to synthesize specific pigments. The evolution of the developmental mechanisms that establish focus was therefore the key to the origin of butterfly eyespots. [Carroll *et al.*, 1999, 532]

What Carroll's team discovered is that the genes and the entire regulatory pathway that integrates them and which control anterior/posterior wing development in the *Drosophila* (or its common ancestor with butterflies) have been recruited and modified to develop the eyespot focus. This discovery of the "facility with which new developmental functions can evolve. . . within extant structures" [p. 534] would have been impossible without the successful why-necessary answer to the proximate question of developmental biology.

Besides the genes noted above, there is another, *Hedgehog* whose expression is of particular importance in the initial division of the *Drosophila* wing imaginal disk into anterior and posterior segments. As in the fruit fly, in *Precis* the *hedgehog* gene is expressed in all cells of the posterior compartment of the wing. but its rate of expression is even higher in the cells that surround the foci of the eyespot. In *Drosophila*, *Hedgehog*'s control over anterior/posterior differentiation appears to be the result of a feed back system at the anterior/posterior boundary involving four other gene products, and in particular one, *engrailed*, which represses another, *cubitus interruptus* (hereafter '*ci*' for short), in the fruit fly's posterior compartment. This same feed back look is to be found in the butterfly wing posterior compartment, except that here the *engrailed* gene's products do not repress *ci* expression in the anterior compartment of the wing. The expression of *engrailed*'s and *Ci*'s gene-products together result in the development of the focus of the eyespot.. One piece of evidence that switching on the *Hedgehog-engrailed-ci* gene system produces the eyespot comes from the discovery that in those few butterflies with eyespots in the anterior wing compartment, *engrailed* and *ci* are also expressed in the anterior compartment at the eyespot foci (but not elsewhere in the anterior compartment). "Thus, the expression of the *Hedgehog* signaling pathway and *engrailed* is associated with the development of all eyespot foci and has become independent of the [anterior/posterior] restrictions [that are found in *Drosophila*]" [p. 534]

Further experiments and comparative analysis enabled Carroll and co-workers to elucidate the causal order of the changes in the *Hedgehog* pathway as it shifts from

wing-production in *Drosophila* (or its ancestor) to focus production in *Precis* eyespot development. "The similarly between the induction of *engrailed* by *Hedgehog* at the [anterior/posterior] boundary [of both fruit fly and butterfly wings, where it produces the intervein tissue in wings] and in eyespot development suggests that during eyespot evolution, the *Hedgehog*-dependent regulatory circuit that establishes foci was recruited from the circuit that acts along the Anterior/Posterior boundary of the wing." [p. 534].

Of course, the full why-necessary proximate explanation for any particular butterfly's eyespots is not yet in, nor is the full why-necessary proximate explanation for the development of the *Drosophila's* (or its ancestor's) wing. But once they are in, the transformation of the ultimate explanation of why butterflies have eyespots on their wings into a proximate explanation can begin. This fuller explanation will still rely on natural selection. But it will be one in which the alternative available strategies are understood and the constraints specified, the time and place and nature of mutations narrowed, in which adaptations are unarguably identifiable properties of genes–their immediate or mediate gene products (in Dawkin's terms, their extended phenotypes), and in which the feedback loops and causal chains will be fully detailed, and the scope for doubt, skepticism, questions and methodological critique that ultimate explanations are open to will be much reduced.

The macromolecular reductionist holds that why-necessary explanations can only be provided for by adverting to the macromolecular states, processes, events, and patterns that these non-molecular historical events and patterns supervene on. Any explanation that does not do so, cannot claim to be an adequate, complete why-necessary explanation. The reductionist does not claim that biological research or the explanations it eventuates in can dispense with functional language or adaptationalism. Much of the vocabulary of molecular biology is thoroughly functional. Nor is reductionism the claim that all research in biology must be "bottom up" instead of "top down" research. So far from advocating the absurd notion that molecular biology can give us all of biology, the reductionist's thesis is that we need to identify the patterns at higher levels because they are the explananda that molecular biology provides the explanantia for. What the reductionist asserts is that functional biology's explanantia are always molecular biology's explananda.

So, why isn't everyone a reductionist, why indeed, does antireductionism remain the ruling orthodoxy among philosophers of biology and even among biologists? Because, in the words of one antireductionist, reductionism's alleged "mistake consists in the loss of understanding through immersion in detail, with concomitant failure to represent generalities that are important to 'growth and form'. [Kitcher, 206]. The reductionist rejects the claim that there is a loss of biological understanding in satisfying reductionism's demands on explanation, and denies that there are real generalities to be represented or explained. In biology there is only natural history — the product of the laws of natural selection operating on macromolecular initial conditions.

Reductionism accepts that selection obtains at higher levels, and that even for some predictive purposes, focus on these levels often suffices. But the reductionist

insists that the genes, and proteins they produce are still the "bottleneck" through which selection among other vehicles is channeled. Without them, there is no way to improve on the limited explanatory power to be found in functional biology. Insofar as science seeks more complete explanation for historical events and patterns on this planet, with greater prospects for predictive precision, it needs to pursue a reductionistic research program. That is, biology can nowhere remain satisfied with how-possibly ultimate explanations, it must seek why-necessary proximate explanations, and it must seek these explanations in the interaction of macromolecules.

But this argument leaves a hostage to fortune for reductionism about biology, one large enough to drive home a decisive antireductionist objection. Although the reductionism here defended claims to show that the how-possible ultimate explanations must be cashed in for why-necessary ultimate explanations, these explanations are still ultimate, still evolutionary — they still invoke the principle of natural selection. And until this principle can be reduced physical law show, it remains open to say that even at the level of the macromolecules, biology remains independent from physical science. Thus, the reduction of molecular biology to physical science remains an agenda item for physicalism

BIBLIOGRAPHY

[Carroll *et al.*, 1999] S. Carroll, *et al.*. Recruitment of a *Hedgehog* regulatory circuit in Butterfly eyespot evolution, *Science*, 283:532-534, 1999.

[Feyerabend, 1964] P. Feyerabend. Reduction, Empiricism and Laws, *Minnesota Studies in the Philosophy of Science, v. III*, Minneapolis, University of Minnesota Press, 1964.

[Garfinkel, 1981] A. Garfinkel. *Forms of Explanation*, New Have, Yale University Press, 1981.

[Gould and Lewontin, 1979] S. J. Gould and R. Lewontin. The spandrels of St. Marco and the Panglossian paradigm, *Proceedings of the Royal Society of London*, B205: 581-598 1979.

[Kim, 1992] J. Kim. Downward Causation. In A. Beckermann, H. Flohr and J. Kim, eds. *Emergence or Reduction?*, pp. 119-138. Berlin: de Gruyter, 1992.

[Kitcher, 1984] P. Kitcher. 1953 and all that: a tale of two sciences, *Philosophical Review*, 93, pp. 353-373, 1984.

[Kitcher, 1999] P. Kitcher. The hegemony of molecular biology, *Biology and Philosophy*, 14: 195-210, 1999.

[Kuhn, 1962] S. Kuhn. *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press, 1962

[Mayr, 1982] E. Mayr. *The Growth of Biological Thought*, Cambridge, Belnap Press, 1982

[Nijhout, 1994] F. H. Nijhout. Genes on the Wing, *Science*, 265: 44-45, 1994

[Putnam, 1975] H. Putnam. *Mind, Language, and Reality* , New York: Cambridge University Press, 1975

[Rosenberg, 1985] A. Rosenberg. *Structure of Biological Science*, Cambridge, Cambridge University Press, 1985.

[Rosenberg, 2006] A. Rosenberg. Darwinian Reduction or How to Stop Worrying and Love Molecular Biology. University of Chicago Press, 2006.

[Schaffner, 1967] K. Schaffner. Approaches to Reductionism", *Philosophy of Science*, 34: 137-147, 1967.

[Wolpert *et al.*, 1998] L. Wolpert, *et al.*. *Principles of Development*, Oxford, Oxford University Press, 1998.

TRAITS, GENES, AND CODING

Michael Wheeler

1 THE UNIQUENESS OF GENES

Although, in most biological circles, talk of the causes of phenotypic structure naturally invites talk of genes, *everyone* knows (or ought to) that biological development is a mightily complex process involving a vast array of causal factors, some of which are genetic and some of which aren't. Elements with developmentally decisive effects are easily discoverable in non-genetic constituencies such as the gene's surrounding metabolic context and the developing organism's environment (examples below). Thus *everyone* knows (or ought to) that genetic and non-genetic factors *interact* during development, thereby causally *combining* to produce the phenotype. Such is the received wisdom in contemporary biological thought. Nevertheless, and in spite of this *interactionist consensus* (a term I have borrowed from [Sterelny and Griffiths, 1999]), the fact is that among all the co-contributing developmental factors, genes remain special. That, anyway, is what we're told. So what mandates this prioritizing of the gene?

This is the point at which the concept of genetic coding makes its entrance onto the theoretical stage. The view that genes, or complexes of genes, code for phenotypic traits is just as much a part of the current biological orthodoxy as the interactionist account of development, and goes hand in glove with it. Such coding talk, which is of course a species of *representational* explanation, is, if not ubiquitous, overwhelmingly common, both within the scientific community and beyond. Indeed, it is the keystone of popular views according to which the genotype as a whole should be conceived as a set of *instructions* for, a *blueprint* for, a *plan* for, a *specification* of, or a *program* for, the building of the phenotype. All the highlighted notions, while perhaps subject to subtle differences in meaning that might be important in particular contexts, depend conceptually on the idea that genes make a representational contribution to development. The idea that genes code for phenotypic traits is thus an ineliminable component of such views. Moreover, one conceptual stage back, representation-talk gets a grip only where it makes explanatory sense to think in terms of structures that carry, are vehicles for, exploit, or in some other way trade in, *information*. Whether or not one can think of structures as information-carrying (in a rich semantic sense, rather than merely in information-theoretic terms — see below) without thereby thinking of those structures as representational is, I think, a moot point. In any case I shall take it that representation-talk requires information-talk, so establishing that the

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

latter makes sense is a significant step towards establishing that the former does too.

Among other things, coding talk about genes is supposed to help us make good on the claim that genes are special developmental factors, that they count as being *privileged* causal elements in the developmental process. The way that coding talk is supposed to achieve this feat is nicely captured by Lorenz's [1965] image of the non-genetic causal factors in development as nothing more than the building blocks out of which organisms are systematically constructed according to a blueprint stored in the genes. On such a view, the real challenge for developmental biology is to understand how genetically specified instructions organise those available developmental materials into an organism. This way of looking at things really does make genes special.

There are many gene concepts in the literature, ranging from the essentially abstract, generically Mendelian notion of a gene as a trait difference marker to various attempts to give molecular substance to the idea.[1] There are even some who argue that most of the assumptions that historically have underpinned the term 'gene' have been shown to be problematic, meaning that the very concept of a gene is now, in many ways, a misleading one that perhaps biology could do without (e.g., [Dupre, 2005]). For the purposes of the present investigation I intend to put both definitional diversity and strategic critical eliminativism aside, and stipulate that we should be thinking of a gene as an entity with some sort of molecular unity, that is, as a stretch of DNA that possesses some sort of ontological integrity. To make this idea firm enough for the job at hand, we need to resist the tempting thought that the way to establish the molecular unity in question is by holding that genes are those parts of the genome that code.[2] Why this is should be clear enough: I have been assuming that there is conceptual space for the following result: there are genes but they don't code for anything. If genes simply are the coding parts of the genome, then this result is not available. A negative answer to the question 'do genes code?' would imply that there are no genes. So we need to achieve the desired molecular unity without appealing to coding. But how? One answer would be to appeal to causally underpinned structural isomorphisms that exist between (a) sequences of DNA and (b) certain developmental elements that are causally downstream of those sequences. The most likely candidates for the latter are proteins, since the claim that systematic causal mappings exist between sequences of DNA and amino acids in proteins is not generally thought to be controversial (see the description of protein synthesis in section 4 below). Of course, if (i) all there is to coding is some sort of systematic causal dependence, and (ii) genes may rightly be said to code for proteins, whether or not they also code for traits, then the recognition of systematic causal mappings between (a) and (b) would herald significant progress in our investigation. Genes would code

[1] For a recent review, see Griffiths and Stotz forthcoming. See also the *Representing Genes* project at *http://www.pitt.edu/∼kstotz/genes/genes.html*.

[2] Contrary to some accounts, an organism's genome is not simply its complete set of genes, but much more besides (see [Dupre, 2005]).

for proteins (at least). However (i) is implausible, as we shall see in section 3, so even if (ii) is true, it can't be on the basis of (i). The upshot is that, at a programmatic level, we are in a position to identify genes in advance of settling the coding issue.

If the primary goal of introducing the concept of genetic coding is to single out genes as privileged causal elements in the developmental process, then it might well seem that any successful account of coding talk must have the consequence that, of the many causal factors that combine causally during development, it is the genes *alone* that end up coding for phenotypic traits. Let's call this *the uniqueness constraint*. (Griffiths and Knight [1998] introduce what is essentially this very constraint in terms of what they call the 'parity thesis'; see also [Griffiths, 2001].) The uniqueness constraint will not be met if either (a) the account of genetic coding under consideration fails to deliver the result that genes code for traits, since if genes don't code for traits then they can't do so uniquely, or (b) that account does deliver the result that genes code for traits, but its conditions for what it is to do this are met by other elements in the extended developmental system, since then genes won't be the only developmental elements that code for traits.

It's an irritating but undeniable fact that the natural world rarely plays ball with neat philosophical distinctions and categories, so the uniqueness constraint, in the strict form just stated, is very likely to be violated by any non-question-begging account of genetic coding on which we settle. Still, as long as such violations are not the norm, they are of no great matter. The background methodological thought concerning the genetic target of coding talk in biology can surely tolerate the odd non-genetic interloper. To be sensitive to this state of affairs we can modify the uniqueness constraint slightly, to require only the following: any successful account of genetic coding must have the consequence that those non-genetic elements for which it would be unreasonable, extravagant, or explanatorily inefficacious to claim that their contribution to development is representational in character do not count as coding for developmental outcomes. Call this the *weakened uniqueness constraint*. The weakened uniqueness constraint still has teeth, since the overwhelming majority of non-genetic developmental factors surely belong in the non-representational category. So perhaps it's acceptable for, say, an antero-posterior gradient of the bicoid protein in the *Drosophila* egg to be a vehicle of representational content (see [Maynard Smith, 2000b]), but not, say, environmental temperature or the force of gravity (see below). There will no doubt be borderline cases to be fought over. Let's use the term *illegitimate non-genetic elements* to label those non-genetic factors for which it would be unreasonable, extravagant, or explanatorily inefficacious to claim that they code for developmental outcomes. So *legitimate non-genetic elements* are those non-genetic elements for which it would be reasonable, prudent, and explanatorily efficacious to claim that they code for developmental outcomes. We can now state an important principle: if one is considering the proposal that meeting certain specified conditions is sufficient for representing phenotypic structure, and it turns out that adopting those

conditions would allow not only genes and legitimate non-genetic elements, but also illegitimate non-genetic elements to qualify (that is, there is a transgression of the weakened uniqueness constraint), then one should conclude that the proposed conditions are in fact *not* sufficient for representation.[3]

Using the benchmark of meeting the weakened uniqueness constraint as a sign of success, is it possible to give an adequate account of genetic coding? What follows is an attempt to answer this question. I should warn you that it won't exactly be a stroll in the park. Here's the route: Having set things up by saying more about exactly why the massively distributed character of the causal systems underlying development might actually be in tension with coding talk about genes (section 2), I shall consider the main contenders from the literature that purport to be not only plausible reconstructions of the character of such talk, but also justifications of its explanatory efficacy, and I shall find each of them wanting (section 3). At that point in the proceedings I shall lay out an alternative and, I suggest, superior strategy for understanding and justifying coding talk in the relevant area of biology (section 4), but argue that that strategy has at least one quite radical implication that is, I think, a bullet that we just have to bite (section 5). In the final section (section 6), I shall consider an objection to the claim that there is coding for traits, an objection that applies to all the candidate strategies on the table, including the one I favour.

2 CAUSE FOR CONCERN

In recent years some of the most persistent critics of the idea that genes are informational entities that code for traits have come from the ranks of the developmental systems theorists. (For classic statements of the developmental systems position, sometimes just called developmentalism, see, e.g., [Oyama, 1985]; [Griffiths and Gray, 1994]; [Griffiths and Knight, 1998]; and various papers in [Oyama *et al.*, 2001].) Developmental systems theorists hold that the fundamental unit of evolution is the life cycle (a process that reconstructs itself from one generation to the next using a suite of developmental resources). Given that they take the life cycle to be the basic evolutionary unit, developmental systems theorists object to any view that understands development in terms of some basic dichotomy between genes and the rest of the extended developmental system. Thus they reject (what they see as) the massive over-emphasis on genes in (what they see as) mainstream neo-Darwinian evolutionary biology. It is important to be clear here

[3]See [Wheeler, 2005, 208–209], for similar moves in the case of the neural target of the concept of representation in cognitive science. For what I take to be a similar weakening of (what I am calling) the uniqueness constraint in the case of genetic coding, see [Stegmann, 2005]. To keep a sense of balance, it is worth noting that Sarkar [2000; 2005] explicitly recommends that a constraint which is closely analogous to the uniqueness constraint be dropped (at least for eukaryotes), on the grounds that no conceptually respectable concept of genetic information is available which doesn't have the consequence that that constraint is violated. I suggest, by contrast, that any notion of genetic information which has the consequence that the (weakened) uniqueness constraint is violated thereby loses its claim to conceptual respectability.

that developmental systems theorists are not denying that there are any interesting empirical differences between the ways in which, say, DNA sequences and, say, parental scaffolding of language learning during early childhood contribute to development. What they deny is that these empirical differences should be turned into what Griffiths [2001, 406] calls a "scientific metaphysics." As Griffiths and Gray put the point:

> [G]enes are just one resource that is available to the developmental process. There is a fundamental symmetry between the role of the genes and that of the maternal cytoplasm, or of childhood exposure to language. The full range of developmental resources represents a complex system that is replicated in development. There is much to be said about the different roles of different resources. But there is nothing that divides the resources into two fundamental kinds. The role of the genes is no more unique than the role of many other factors. [Griffiths and Gray, 1994, 277–304]

One sure-fire route to the sort of scientific metaphysics that developmental systems theorists reject would be to adopt coding talk about genes alongside the uniqueness constraint (in either its full-strength or its weakened form), and to suggest that (all or the vast majority of) non-genetic developmental factors should, in a Lorenzian fashion, be relegated to mere genetically assembled building blocks. With this line of thought in their critical sights, Griffiths and Knight [1998] claim that "DNA does not contain a program for development" (p. 253) and deny that there are "pre-formed blueprints or representations of traits in DNA" (p. 255).

This is not the place to become over-focused on the details of the developmentalist agenda. Our concern will be with a general way of motivating anti-representationalism about genes that is often at work in developmental systems thinking, as well as in the arguments of other prominent genetic coding sceptics who lay stress on the distributed character of the causal processes underlying development (for example, [Maturana and Varela, 1987], more on whom below). To bring things into focus, it will be useful to highlight a phenomenon that Andy Clark and I have dubbed *causal spread* ([Wheeler and Clark, 1999]; see also [Wheeler, 2003; 2005]). Causal spread obtains when some phenomenon of interest turns out to depend upon causal factors external to the system previously or intuitively thought responsible. Thus the identification of causal spread depends on the previously accepted explanation of the phenomenon of interest. Of course, given some default view of the world, even the most mundane examples of representational systems might display some degree of causal spread. For example, we might reasonably think of a C program as a set of instructions for (i.e., as a set of representations of) computational outcomes. The fact is that a C program is nigh on useless without certain 'environmental' (with respect to the program) features, such as a working operating system. However, nothing about the positive representational status of the C program would be threatened by the discovery of the essential causal contribution of the operating system.

Having said that, not all modes of causal spread are quite so obviously harmless to representational explanation. Consider what one might call *non-trivial causal spread*. This phenomenon arises when the newly discovered additional causal factors reveal themselves to be at the root of some distinctive target feature of the phenomenon of interest. In effect, where one confronts non-trivial causal spread, a new sharing-out of the explanatory weight is mandated. Call this *explanatory spread* [Wheeler and Clark, 1999]. Mameli [2005] explains the key points in this way.

> Causal spread occurs when we discover some new factor causally involved in the occurrence of a phenomenon. Explanatory spread occurs when we realize that some factor that was not considered to be necessary in the explanation of a phenomenon is instead explanatorily necessary for that phenomenon. Or, to put it differently, explanatory spread occurs when we realize that some factor that was not taken to be part of a sufficient explanation of a phenomenon needs to be included in such explanation. Since the fact that something is causally required does not entail that it is also explanatorily required, causal spread does not necessarily lead to explanatory spread. But in cases where the newly discovered causal factor is deemed to be an important one, causal spread is likely to generate the inclusion of the newly discovered factor in any sufficient explanation of a phenomenon to which this factor causally contributes. That is, in these cases, causal spread leads to explanatory spread. [Mameli, 2005, 388]

In the present paper, the phenomenon of interest is organismic structure, and the default position is that such structure is down to genetic coding (on something like a Lorenzian model according to which the non-genetic material causes in development are the bricks and mortar out of which the organism is assembled according to the genetic blueprint). Against this background, one would have non-trivial causal spread where one discovered a distributed developmental system in which non-genetic organismic and/or wider environmental factors made explanatorily non-negligible contributions to phenotypic form. So, is there non-trivial causal spread, and thus explanatory spread, in (our theories of) biological development? The answer, surely, is yes. Developmental explanatory spread is common. I shall give just a few brief illustrative examples, but the biological literature is simply brimming over with others.

First, consider the process of determination during cell specialization. In vertebrates, prior to the third cleavage stage, the cells in the developing embryo retain the possibility of achieving any of the full range of developmental outcomes available to the original zygote. The process of determination, in which the future course of development in the cells is differentially restricted, depends on a process in which the nuclei of the various cells become embedded in different cytoplasmic environments which in turn have different regulatory effects on the genes within the various nuclei. The sources of this differential embedding are a range

of non-genetic factors, including pH balance and gravity, which result in a non-homogenous distribution of cytoplasmic materials within the egg. The inclusion of such non-genetic factors in our explanation is thus necessary if we are to account for the phenotypic phenomenon of interest.

Now consider the Mississippi alligator. These creatures lay their eggs in a nest of rotting vegetation which produces heat in varying quantities. Eggs that develop at lower temperatures (within some overall range) end up producing females, whilst those that develop at higher temperatures end up producing males. Eggs in a clutch will pass through the critical developmental window at various different temperatures, meaning that a mixture of females and males will be born. This environmental method of regulating sex ratio (a ratio which, for reasons of population-survival, needs to stay somewhere near 50:50 in the population) might seem a little hit and miss, but it works well enough (for more details, see, e.g., [Goodwin, 1994, 38]). Environmental temperature is a non-genetic factor, the inclusion of which in our explanation is necessary if we are to account for the phenotypic phenomenon of interest.

Finally, turning to human development, there is the much-studied phenomenon of scaffolding, in which a caregiver provides an on-line support system to enable a child to complete a task. As the child displays improving competence at the task, the caregiver gradually withdraws the support system, transferring responsibility for the completion of the task to the child (see, e.g., [Wood *et al.*, 1976]). Scaffolding is a key feature of child development, in areas such as discourse participation, literacy, and self-regulation, although the style and extent of the caregiver intervention varies among cultures. Scaffolding is a non-genetic factor, the inclusion of which in our explanations is necessary if we are to account for a range of phenotypic phenomena of interest.

Taking the foregoing examples as paradigmatic of development, we can conclude that explanatory spread is rife in that arena. What we can't conclude right now is that this generates a problem for coding talk about genes. For even though a new sharing out of the explanatory weight is mandated, such that non-genetic elements such as pH balance, gravity, environmental temperature and caregiver scaffolding become part of the relevant explanatory matrix, we haven't yet found out exactly why that fact might undermine the positive representational status of the genetic contribution.[4] So let's turn now to an explicit argument against the view that genes code for phenotypic traits, one that appeals to (what I am calling) developmental explanatory spread.

> We have often heard it said that genes contain the "information" that specifies a living being ... [but] when we say that DNA contains what

[4]To be clear: it is highly plausible that the kinds of non-genetic factors highlighted are illegitimate non-genetic factors, in the sense that *if* they counted as representations of developmental outcomes according to some account of what it is for an element to play that role, then that in itself would be grounds for rejecting the proposed account, since the weakened uniqueness constraint would have been violated. But we don't as yet have such an account on the table. Our investigation hasn't progressed that far.

> is necessary to specify a living being, we divest these components ... of their interrelation with the rest of the network. It is the network of interactions in its entirety that constitutes and specifies the characteristics of a particular cell, and not one of its components. That modifications in the components called genes dramatically affect the structure is very certain. The error lies in confusing essential participation with unique responsibility. By the same token one could say that the political constitution of a country determines its history. This is obviously absurd. The political constitution is an essential component in any history but it does not contain the "information" that specifies that history. [Maturana and Varela, 1987, 69]

Maturana and Varela's claim is that the fan of genetic information mistakenly confuses "essential participation with unique responsibility". This suggests that for genes to count as carrying the information that specifies phenotypic traits, and thus for genes to be in the right conceptual ballpark to code for such traits, genes would need to bear *sole responsibility* for phenotypic form. But if, as the examples discussed earlier suggest, biological development is a playground for explanatory spread, then any such description of the genetic contribution here looks to be unwarranted. In general, DNA will *not* meet the sole responsibility condition. So it seems that if the representational theory of genes is tied to this condition, then that theory is straightforwardly undermined by the presence of developmental explanatory spread. And that, in essence, is Maturana and Varela's point when they say, with respect to the cell, that it is "the network of interactions in its entirety that constitutes and specifies the characteristics of a particular cell, and not one of its components".

But now surely something has gone wrong. Given my opening remark that every biologist understands (or ought to understand) development as involving a vast range of genetic *and non-genetic* causal factors, Maturana and Varela's argument seems to do no more than set up a straw man for summary execution. However, things are not quite that simple. Indeed, despite the pretty much universal acknowledgement that there are extra-genetic causal contributions to development, the fact is that many theorists fall prey to the following, seductive thought: if one could find out the complete sequence of an organism's DNA, then, in principle, one would be able to use that information *alone* to compute the adult organism, such that one would be able to predict, in every relevant detail, that adult's phenotypic form. As DeLisi puts it:

> The collection of chromosomes in the fertilized egg constitutes the complete set of instructions for development, determining the timing and details of the formation of the heart, the central nervous system, the immune system, and every other organ and tissue required for life. [DeLisi, 1988, 488]

At work here is a deceptively tempting view of outcome-directed representation that Clark and I have previously dubbed *strong instructionism* ([Wheeler and

Clark, 1999]; see also [Wheeler, 2003; 2005]). Strong instructionism is the claim that what it means for some element to code for an outcome is for that element to fully specify the distinctive features of that outcome, where 'full specification' requires that the kind of exhaustive predictive power just indicated may, in principle, be achieved on the basis purely of what may be known about the putatively representational factor. In the present context, strong instructionism amounts to the claim that what it means for a gene (or a complex of genes) to code for a phenotypic trait is for that gene (or complex of genes) to fully specify the form of that trait. (Here we finally see the true colours of that compelling Lorenzian image of blueprints and materials.) However, given the presence of developmental explanatory spread, the fact is that knowing the entire sequence of an organism's DNA will *not* be sufficient to predict phenotypic form. So it seems that if coding talk about genes is tied to strong instructionism, then such talk is unsustainable.

Still, when it comes to providing a satisfactory account of genetic coding, there's *something* right about strong instructionism, namely that it respects the following, eminently plausible principle: in counting some target factor as a representation, in an appropriate outcome-directed sense, one buys into a crucial asymmetry between, on the one hand, that putatively representational factor and, on the other, the ecological backdrop against which that factor operates. Indeed, in all cases of algorithms, programs, instruction-sets, and other action-producing codes, those representational states and processes are able to perform their outcome-generating functions only given some assumed backdrop of other causally active states and processes. To build on a previous example: try running a C program without certain 'environmental' (with respect to the program) features, such as a working operating system. Moreover, where the right kind of asymmetry exists in the extended causal system, the discovery of causal spread, *even of the non-trivial variety that generates explanatory spread*, will not undermine representationalism. Thus we may conclude that it will be legitimate to treat genes as coding for traits, even in the face of developmental explanatory spread, just so long as we can legitimately regard the rest of the extended developmental system as the ecological backdrop against which genes make their representational contributions to phenotypic outcomes.

Notice that nothing about this suggestion requires that the crucial asymmetry be established independently of whatever detailed account we give of genetic coding. Rather, an adequate account of genetic coding should have the consequence that the right kind of asymmetry is manifest. We can now see how our overall benchmark for success, meeting the weakened uniqueness constraint, fits into the current dialectic. As I argued earlier, any satisfactory account of the concept of genetic coding must have the following consequences: (a) if any non-genetic factors count as coding for traits, then such violations of the uniqueness of genes in being representations of developmental outcomes should not be the norm; and (b) where such violations do occur, it should be neither unreasonable, nor extravagant, nor explanatorily inefficacious to claim that the developmental contribution of the non-genetic factors in question is representational in character. In singling out genes

as the predominant causal elements in the extended developmental system that code for traits, we simultaneously earn the right to treat the rest of that system as an ecological backdrop against which those genes (along with perhaps certain legitimate non-genetic elements) operate. Strong instructionism meets this demand through the full specification condition and the associated Lorenzian claim that non-genetic developmental factors in general are no more than biological bricks and mortar. But this view of non-genetic factors is not available once developmental explanatory spread is in the picture. So we are left with a challenge. What we need is an account of genetic coding that, without imposing the full specification condition, meets the weakened uniqueness constraint. In the next section I discuss a number of (ultimately unsuccessful) ways of addressing this challenge.

3 FALSE STARTS AND DEAD ENDS

Here's a seductive first shot: genes code for traits because they *causally co-vary* with traits. In other words, appropriate causal co-variation is sufficient for genetic representation. One reason why this suggestion is provisionally attractive is that it makes contact with well-established views from elsewhere in science and philosophy that treat information in purely causal terms, or at least that might be used to explicate such an idea. Thus, at a first pass, causal information might, in part, be cashed out by way of mathematical information theory [Shannon and Weaver, 1949], according to which (roughly) the quantity of information in a system is identified with the amount of order in that system. I say 'in part' because, strictly speaking, Shannon information supposes only correlation rather than causal correlation, so the causal nature of the correlation is an extra feature. I say 'at a first pass' because, for the purposes of genetic information, where we mostly want to talk about the *content* of the information in a system, rather than how much of it there is, the notion of causal information is more usefully explicated in the light of Dretske's [1981] influential philosophical treatment. Here is the resulting picture. Where there exists a sending system and a receiving system, connected by a channel such that the state of one system is causally related, in a systematic way, to the state of the other, then we have a signal — a flow of information — between the two systems. The causal information content of the signal is the source with which it is reliably correlated. This account is straightforwardly adapted such that entities carry information about causally downstream states with which they co-vary.

So how useful are causal information concepts in the present context? Mahner and Bunge [1997] question their applicability. First they point to the largely noiseless character of the (so-called) genetic code, noting that, practically speaking, the presence of noise is a standard issue when deploying Shannon information. Second, they claim that chemical processes cannot be thought of as signals that carry messages. In response, Maynard Smith [1999] argues (rightly in my view) that typesetting is largely noiseless, yet causal information concepts would surely be applicable there, and that it's hard to see why chemical processes couldn't be

vehicles of causal information content, since all manner of other physical media, such as fluctuating currents in wires and sound waves, are standardly thought to be good for the job. A more serious barrier to the use of causal information concepts in genetics is that, given the standard conception of the genome as specifying phenotypic outcomes in a disjunctive manner (i.e., develop like *this* under *these* environmental circumstances, like *this* under *these* environmental circumstances, and so on), the causal information view licences us to speak about genetic coding in ways that biologists don't. For example, to use an example due to Griffiths [2001], on the basis of a purely causal notion of information, the human genome would encode the instruction "when exposed to the drug thalidomide grow only rudimentary limbs." But biologists are unlikely to be tempted by such a claim. What this tells us is that the notion of causal information fails to capture the standard usage of informational terms in biology.

The most substantial problem confronting claims that appropriate causal co-variation is sufficient for genetic representation (or for genetic information), however, is one of *excessive liberality*. It is indeed a familiar point from the literature that genes are not the only factors in the developmental system that might be identified as causally co-varying with traits. Of course, it seems clear enough that if one could hold non-genetic causal factors in the developing body and the environment constant, while varying the genotype, then one would find causal co-variations between genes and phenotypic traits. However, if one could hold the genotype and the non-genetic causal factors in the developing body constant, while varying environmental factors, then one would find causal co-variations between environmental variables and phenotypic traits. Similarly, if one could hold the genotype and the environment constant, while varying non-genetic causal factors in the developing body, then one would find causal co-variations between those factors and phenotypic traits. But now if causal co-variation is a sufficient condition for a developmental factor to be representational, and if non-genetic causal factors in the developing body and the environment can causally co-vary with phenotypic traits, then those extra-genetic elements will sometimes count as coding for traits. This spells trouble because, given that many of the non-genetic factors here will be illegitimate ones, it falls foul of the weakened uniqueness constraint. In short, as a sufficient condition for coding, causal co-variation is excessively liberal, in that it licences explanations in which *too much* of some extended developmental system might emerge as coding for traits. So while it is eminently plausible that appropriate causal co-variation is necessary for genetic representation, it cannot be sufficient; genetic representation must be appropriate causal co-variation *plus something else*.[5]

What might that something else be? Here is a suggestion: genes code for traits because they (additionally) *set certain parameters* for the developmental systems that generate phenotypes. Perhaps then we can say that while genes do not fully

[5]The fact that any systematic causal co-variation account of genetic coding will be excessively liberal (in the sense identified in the main text) is widely appreciated; see, e.g., [Griffiths and Gray, 1994]; [Maynard Smith, 2000a]; [Griffiths, 2001]; [Sarkar, 2005].

specify the final phenotypic form (strong instructionism is false), they nevertheless code for developmental parameters and, by extension, for phenotypic traits (see [Maynard Smith, 1998] for a version of something like this strategy). The claim that genes might broadly be conceptualized as setting developmental parameters ought not, I think, to be particularly controversial. As Goodwin [1994, 102] puts it, "[d]uring reproduction, each species produces gametes with genes defining parameters that specify what morphogenetic trajectory the zygote will follow". However, it is quite another matter to claim that developmental parameter-setting is sufficient for representation, in the relevant sense. Indeed, this idea suffers from a version of the very excessive liberality problem that dogged the causal co-variation proposal. There seems little doubt that certain non-genetic factors (e.g., environmental temperature in the case of sex determination in the Mississippi alligator) might, like genes, be treated as parameterizing developmental systems. These non-genetic factors would then co-specify, along with the relevant genes, exactly which possible trajectory of that system would finally be traversed by the developing organism. But if performing the function of parameter-setting is sufficient for some developmental factor to count as coding for a phenotypic trait, then these extra-genetic factors will qualify. And, given that many of the non-genetic factors here will be illegitimate ones (environmental temperature would be an example), that violates the weakened uniqueness constraint. So even if performing the function of setting developmental parameters is necessary for a causal factor to play a representational role in development (which it might be, if one conceives of developmental systems as dynamical systems), such a role cannot be *just* a matter of developmental parameter-setting; it must be developmental parameter-setting *plus something else*.

It is time for a tactical rethink. So far we have considered, only to reject, two versions of the view that the status of genes as coding for traits is secured by properties of the direct causal contribution of genes. Perhaps the problem is that we're looking at things all wrong. Perhaps representation is a matter of *function* rather than (brute) causation. In evolutionary biology, function-talk naturally invites an appeal to Darwinian selection. On this view, the function of a developmental element (if it has one) is (roughly) the positive contribution to organismic survival and reproduction prospects that ancestors of that element have made within historical populations. This generates the following proposal: genes code for traits insofar as they have been *selected* precisely so that a particular trait should occur (see, e.g., [Sterelny, 1995]; [Maynard Smith, 2000a]).

Why might someone think that appealing to selection is a good way to go on the issue of genetic coding? One motivating thought is that the concept of information that matters to biology is not causal information, but *intentional (or semantic) information*.[6] The intentional concept of information is modelled on

[6]For this distinction drawn in these terms, see, e.g., [Sterelny and Griffiths, 1999]; [Maynard Smith, 2000a]; [Griffiths, 2001]. For scepticism about the applicability, within the genetic context, of the intentional concept of information, see [Sarkar, 2005]. Sarkar presents his own account of genetic information, in terms of what he calls 'semiotic information,' which, while being

the kind of information carried by human thoughts and utterances. And one of the standard philosophical tests for the presence of intentional information is to see if one can make sense of the phenomenon of misrepresentation. In cases of input-related mental representations, misrepresentation occurs when the content of the representational state fails to correspond to the state of affairs in the world that caused it (e.g., one's 'cow' representation is activated by perceptual contact with a horse). In cases of outcome-directed mental coding, misrepresentation occurs when the content of the representational state fails to correspond to the state of affairs that it helps bring about (one's grasp-controlling representations are activated but, due to intervening causes, fail to result in the beer glass leaving the table). Genetic coding, if it exists, is, of course, an outcome-directed form of representation. So misrepresentation would occur if the content carried by the gene (its developmental instruction) fails to correspond to the phenotypic state of affairs that it helps bring about (the gene coding for long legs is causally active but, due to intervening causes in the developmental system, the phenotype ends up with short legs).

In outcome-directed mental representation, misrepresentation is made possible because the content of the mentally represented action-oriented instructions remains the same, no matter what happens in the rest of the action-generating system. So, in the case of genetic coding, we need it to be the case that the content of the represented instructions remains the same, no matter what happens in the rest of the developmental system. A dramatic illustration of the intuitive plausibility of cross-context content within biological systems comes from some striking experiments due to Halder, Callaerts, and Gehring [1995]. There is a particular gene that plays a causal role in eye development in the mouse. Transfer that gene to the fruitfly *Drosophila* and it will result in the development of an eye — a compound eye, a fruitfly eye. Indeed, activate the transplanted gene at various sites and one will get a fruitfly eye developing at the different organismic locations in question (e.g., at the usual site of a leg). So, if this gene codes an instruction, the content of that instruction is very plausibly something like 'build me an eye here'. That's the developmental instruction represented by that gene.[7] Intuitive plausibility aside, the key point here is that we can make sense of intentional representation because we can make sense of the coding element in question having an 'intended' effect (which in turn determines the content of the represented instructions), even if that effect doesn't come about. Where information is interpreted merely in terms of systematic causal co-variation, there is no room for this distinction between intended and unintended effects, hence the fact that causal information concepts fall prey to the thalidomide counter-example discussed earlier. As Griffiths [2001] notes, the notion of intentional information can handle this case, since growing

deflationary with respect to intentional information, is richer than Shannon information.

[7]My interpretation of this scenario follows that given and defended by Maynard Smith [2000a,b]. For an alternative interpretation, according to which the gene in question should be seen as a reader of information carried by other genes, rather than as carrying information in its own right, see [Sterelny, 2000].

only rudimentary limbs is not one of the intended effects of the genes concerned. But while the idea of the intended effect of a representation might seem straightforward enough in the case of human utterances, exactly how are we to secure that idea in the case of genes? It's here that the appeal to selection comes in. Intended effects are identified by reference to the developmental contribution for which the gene/genes in question was/were selected. So, there is some justification for the claim that an appeal to selection may secure the appropriate sort of informational content for genetic representations. (Whether or not it is the only way to secure such content is another issue — see below.)

Another key thought in the literature is that the appeal to selection will not result in violations of (what I am calling) the weakened uniqueness constraint. Thus Sterelny [1995] observes that the growth patterns of snow gums will differ depending on whether they are exposed to snow or wind. Both genotype and environment are necessary causal factors in determining the plant's final phenotypic form. But whereas the climatic conditions are, in a sense, 'just there', the genotype exists purely because of its role in producing the phenotype, and thus has the evolutionary function of producing the phenotype. And that, according to Sterelny, is why the genotype codes for the phenotype, whereas the environmental factors do not. Notice that, on this view, two genes could play the same brute causal role (say in the production of an eye), but one would rightly be said to code for the relevant property of the eye, while the other wouldn't, if the former had been selected for that job while the latter hadn't.

The suggestion on the table, then, is that genes code for traits insofar as genes, *unlike the rest of the developmental system*, have been selected precisely so that a particular trait should occur. The first thing to say is that selection is not necessary for (genetic) representation ([Sarkar, 2000]; [Wheeler, 2003]). To see why, consider the following argument [Wheeler, 2003]. Genes are sometimes linked physically, in such a way that the evolutionary fate of one gene is bound up with the evolutionary fate of another. This provides the basis for a phenomenon known as genetic hitchhiking. To see how genetic hitchhiking works, let's provisionally allow ourselves the language of 'genes for traits,' and construct a simple evolutionary scenario. Assume that, in some creature, the gene for a thick coat is linked to the gene for blue eyes. Let's also assume that this creature lives in an environment in which it is selectively advantageous to have a thick coat, and selectively neutral to have blue eyes. What will happen is that the gene for a thick coat will be selected for. But since the gene for blue eyes is linked physically to the gene for a thick coat, the gene for blue eyes will be inherited too, even though it bestows no selective advantage, has not been selected for, and thus has no evolutionary function. For present purposes, the key feature of genetic hitchhiking is this: the fact that the hitchhiking gene is not selected for does not in any way threaten, by making theoretically awkward, our description of it as coding for blue eyes. So the phenomenon of genetic hitchhiking tells us that selection is not necessary for representation.

There are two obvious responses that the selectionist about genetic represen-

tation might make. First, she might complain that even if the hitchhiking gene has not been directly selected for, it has a kind of honorary 'selected for' status, on account of the fact that it is linked to a gene that has been directly selected for. But this seems to be the wrong way to describe the situation. After all, in the foregoing example, the gene for blue eyes has certainly not survived because of its role in producing the phenotype. Thus it is hard to see how the notion of being selected for can get any sort of grip. A second response might be to concede that the blue-eyes-related gene does code for blue eyes, but to maintain (i) that selection is sufficient for, but not necessary for, representation, and (ii) that while selection explains why we should describe the thick-coat-related gene as coding for thick coats, some other explanation will be required in the case of the blue-eyes-related gene. But unless there are some powerful independent considerations in favour of clinging on the selectionist strategy (considerations that would have to be produced and judged), there is surely no reason to multiply explanatory stories in this way. What we really want, it seems, is a single account of genetic coding that covers both cases.

Anyway, the fact is that if we adopt the view that selection is sufficient for genetic representation, then, contra Sterelny's snow-gum-driven conclusion, we will fall foul of the weakened uniqueness constraint. To demonstrate this, we can call on a thought experiment due to Mameli [2004]. Consider a species of butterfly with the following properties: (a) all members of the species are genetically identical, and no genetic variation can be produced; (b) the butterflies eat a particular species of plant during the early stages of their life; (c) females lay their eggs on plants of the same species as the one on which they hatch; (d) they do this by eating the leaves of the plant on which they hatch, by imprinting on the taste of the leaves, and by laying their eggs on plants with the same taste. Now, as a result of a developmental accident, the imprinting mechanism in one female malfunctions. She lays her eggs on the 'wrong' plant which, as it happens, is a new species of plant in this species of butterfly's environment. By chance, this new plant makes these butterflies bigger. Now assume that, in this species, bigger size confers a fitness advantage. Because of this, the lucky butterfly's offspring grow up fitter than other butterflies of the species. The offsprings' imprinting mechanisms work just fine. So they lay their eggs on the new species of plant. Given competition for resources, the lucky butterfly's descendants will out-compete their conspecifics and, eventually, all the butterflies of this species will hatch on the new plant. This is a process of natural selection — there is heritable variation in size caused by variation in plant of hatching — but there is no genetic variation. Mameli introduces the term *envirotype* to describe factors such as plant of hatching in the lucky butterfly scenario, factors that are intergenerationally stable (and which thus underwrite selection by guaranteeing a correlation between parental variants and offspring variants), but which are environmental rather than genetic in character. Given the possibility of envirotypes, "not all selection is at bottom genetic selection. Some selection is *nongenetic (or envirotypic)* selection" [Mameli, 2004, 41].

For present purposes, the principal message of the lucky butterfly is that if

being selected for is sufficient for some developmental factor to qualify as coding for a phenotypic trait, then non-genetic factors will sometimes attain coding status, since non-genetic factors may sometimes be selected for. And, of course, if those non-genetic factors are illegitimate ones, as is plausibly the case with plant of hatching in Mameli's thought experiment, then that violates the uniqueness constraint. In short, we confront yet another version of the excessive liberality problem.

It is worth pausing here to note two things. First, the potential existence of Mamelian envirotypes blocks the thought that it must in principle always be possible to trace the adapted character of non-genetic developmental resources back to prior genetic selection (that is, given the suggestion currently on the table, to genes that code for those resources). This is especially clear when the notion of an envirotype is established in the case where genetic variation is ruled out. Second, the conceptual linking of selection to representation, plus the claim that direct selection for non-genetic developmental units is held to be possible, are points embraced by Sterelny and Kitcher in their *extended replicator* proposal [Sterelny and Kitcher, 1988]. According to the idea of extended replicators, *all* adapted developmental resources code for traits. Now, if one interprets the extended replicator proposal as an attempt to reconstruct a theory of *genetic* coding (which is how it seems to be presented by, e.g., [Sterelny and Griffiths, 1999, 87], where it is described as providing a "formal reconstruction of the "gene for" locution"), then one can only assume that Sterelny and Kitcher (a) are unmoved by considerations of the uniqueness of genes with respect to coding status, and (b) do not believe that there is any independent way (independent, that is, of the criterion of selection) to determine whether or not an environmental contribution to development might legitimately qualify as a coding element. My view, as should be clear, is that both (a) and (b) are errors. Of course, it would be entirely consistent to endorse the weakened uniqueness constraint, agree that there are extended replicators, but deny that being selected for is a sufficient condition for coding.

Where next? One intuition that we haven't yet explored is that coding talk is conceptually intertwined with the notion of inheritance. Thus, one might claim that genes code for traits insofar as they are what is passed on from one generation to the next in evolution. Of course, genes *are* inherited. But, using a toy example, let's assume that eye colour can be traced to a single gene, and further that, in a particular offspring, the gene inherited at conception would, if expressed, produce brown eyes. Let's also say that psychology has shown blue eyes to be advantageous to getting on in life, by attracting the favourable attentions of others. This looks like bad news for our target offspring. However, a gene transplant is carried out, such that the inherited brown-eyes-related gene is removed, and a blue-eyes-related replacement inserted by doctors. If we deploy the same style of reasoning as we used in the hitchhiking example above, and provisionally allow ourselves the language of genes as coding for traits, we would naturally say that the inherited, but now removed, brown-eyes-related gene coded for brown eyes. But what about the non-inherited but functional, deliberately inserted, blue-eyes-related gene? As

far as I can see, the fact that this gene has not been inherited does not seem to threaten, or make in any way theoretically awkward, the language of coding. This suggests that being inherited cannot be a necessary condition for coding-talk to get a grip within development.

Moreover, and perhaps more significantly, if we define inheritance without an antecedent pro-gene prejudice, as the biological like-begets-like phenomenon, and so as to fix on elements that are robustly and reliably replicated in each generation of a lineage, and that persist long enough to be the target of cumulative selection, then the fact seems to be that genes are not *all* that organisms inherit. For example, there are so-called *epigenetic inheritance systems*, such as the inheritance of methylation patterns via a separate (from the genetic, that is) copying system; and there is inheritance through *host imprinting*, as when Mameli's imaginary butterflies inherit increased size through imprinting on the taste of a new plant (see above); and then there is the phenomenon of inheritance via *niche construction* [Odling-Smee *et al.*, 2003], as when beaver offspring inherit both the dam that was communally constructed by the previous generation and the altered river flow that that physical structure has produced. Moreover, as Mameli [2005] has argued, simply mentioning DNA-copying and DNA-transmission cannot be sufficient to explain the reliable trans-generational reoccurrence of some phenotypic trait, *if*, that is, one is compelled to mention more than DNA in one's explanation of the development of that trait. Thus:

> If we want to explain why the shape and structure of the legs of human offspring reliably have the same shape and structure as the legs of human parents, we have to mention not only the reliable reoccurrence of the genes involved in normal human leg development, but also the fact that humans experience roughly the same amount of gravitational force from one generation to the next. And this means that, when we explain the reliable reoccurrence ... of legs with a certain structure and shape in human lineages, we have to mention not only DNA-copying and DNA-transmission, but also those processes that explain why human beings experience the same amount of gravitational force generation after generation. [Mameli, 2005, 389]

In short, Mameli's argument is that since there is explanatory spread in (our theory of) development, there is explanatory spread in (our theory of) inheritance.

The upshot of the foregoing observations is that if being inherited is sufficient for some developmental factor to qualify as coding for a phenotypic trait, then non-genetic factors will sometimes count. And if those non-genetic factors are illegitimate ones, as is plausibly the case with the processes that explain why human beings experience the same amount of gravitational force generation after generation, then that once again violates our old friend the weakened uniqueness constraint.

4 A BETTER IDEA

Things are not working out, so let's switch tactics again, and focus our attention on the phenomenon of *protein synthesis*. The guiding intuition here is that something (or some things) about the contribution made by genes to this process will single them out as coding elements, in a way that doesn't contravene the weakened uniqueness constraint.[8]

We should begin by reminding ourselves of some familiar biological facts.[9] In the first stage of protein synthesis, the organism's DNA acts as a template in the manufacture of molecules of *messenger RNA* (*mRNA*). In prokaryotic gene expression, the initial RNA molecule generated by *transcription* (the process underlying templating) is equivalent to the mRNA. However, eukaryotic genes contain sequences of base pairs that are functionally redundant with respect to protein synthesis, sequences known as *introns*. In the initial transcriptional phase, all the DNA (redundant and salient) is transcribed into a complementary RNA copy called *nuclear RNA* (*nRNA*). Then, in a post-transcriptional phase of so-called *RNA splicing*, the introns are subtracted so that only the functionally salient sequences, the *exons*, remain.

The second stage of protein synthesis is known as *translation*. This process is very similar in prokaryotes and eukaryotes, although in prokaryotes transcription and translation are closely coupled, with the latter beginning before the former is complete. In translation, the mRNA molecules produced by transcription (plus RNA splicing in the case of eukaryotes) determine the manufacture of different *proteins*, which are the building blocks of bodies. Molecules of mRNA are divided into triplets of nucleotide molecules known as *codons*, and (ignoring certain singular cases) every instance of a particular mRNA codon, as generated from its DNA template, is believed to result in an instance of the same amino acid being added to an emerging protein. However, this is, as Sarkar [2000, 210] puts it, a "frozen accident". In other words, there is nothing in current biological knowledge to suggest a convincing physical-chemical reason why the mappings could not have been set up differently.[10] So what exactly goes on in translation? The

[8]The thought that the concept of genetic coding will finally be vindicated by facts about the mechanisms of protein synthesis is shared by Wheeler and Clark [1999]; Godfrey-Smith [2000b]; Maynard Smith [2000a]; Sterelny [2000]; Sarkar [2000; 2005]; Wheeler [2003]; and Stegmann [2005]. These alternative developments of the same basic idea contain some significant variations in the precise factors identified as the features of interest, and are occasionally accompanied by certain concessions regarding (a) the full-strength uniqueness constraint and (b) what exactly is represented. Here I shall not attempt to map out *all* the different features that characterize these different views, although it is worth noting at the outset that the concept of arbitrariness (understood one way or another) plays a central role in all of them. The nuances that matter will be discussed as I work towards and defend my own current view.

[9]Protein synthesis is of course a complicated business, and I have no doubt that some readers will be unhappy with one or other aspects of the brief description that I shall give. Nevertheless, the simplified picture I shall paint is broadly correct and good enough for present purposes.

[10]The standard way of describing this frozen accident is to say that the genetic code is arbitrary. As will become clear, however, it is at least plausible that arbitrariness, understood a certain way, is a necessary condition on there being a code *at all*. If that is right, then if the mapping

cell's cytoplasm contains protein-manufacturing-sites called *ribosomes*, along with molecules of another sort of RNA called *transfer RNA (tRNA).* Molecules of tRNA are single nucleotide triplets attached to single amino acids. What happens during translation is that an mRNA molecule becomes attached to a ribosome, and then passes through it, one codon at a time. When a new codon moves into place, the ribosome (through trial and error) locates a molecule of tRNA that, according to the so-called base-pairing rules, features a particular nucleotide triplet. The ribosome then strips off the amino acid from the other end of the tRNA molecule, and adds it to the protein which is under construction. Stripped of its amino acid, the tRNA molecule floats off into the cytoplasm, to be 'recharged' with 'the right' amino acid.

Out of all this biological detail, two conceptually interlocking features of the architecture of protein synthesis strike me as representationally significant.

1. *Arbitrariness*: In the specific sense in which I am using the term, arbitrariness indicates that the equivalence class of different systemic elements (say nucleotide triplets) that could perform some systemic function (say, given other causal factors, produce a specific amino acid) is fixed not by any non-informational physical properties of those elements (say their shape or weight), but rather by their capacity, when organized and exploited in the right way, to carry specific items or bodies of information. The mappings from particular nucleotide triplets to particular amino acids are arbitrary, in this sense.

2. *Homuncularity*: The 'right way' of exploiting the systemic elements just highlighted is established where the system in question is *homuncular.* As I shall use the term, a system is homuncular just when it can be usefully compartmentalized into a set of communicating subsystems, each of which performs a well-defined subtask that contributes towards the collective achievement of a systemic outcome. In an homuncular analysis, the communicating subsystems are conceptualized as trafficking in the information that the inner vehicles carry. More specifically, certain subsystems are interpreted as *producing* information that is then *consumed* downstream by other subsystems. Of course, homuncular subsystems must not be thought of as being, in any *literal* sense, understanders of the information in question. (They are not really little people.) Nevertheless, the fact is that the ways in which the functionally integrated clusters of subsystems exploit inner elements, so as to collectively generate systemic outcomes, become intelligible only if we treat the subsystems involved as dealing in the information that those elements (organized and exploited as they are) carry, rather than as responding only to non-informational physical properties of those elements. The mechanisms underlying protein synthesis are most illuminatingly conceived of as

in question were not arbitrary, there would be no pressure to think of the system in question as one of encodings. So the right question is not "Is the genetic code arbitrary?," but rather, "Is there a genetic code?."

> being homuncular in the requisite sense. Thus mRNA molecules are assembled by a producer subsystem that *encodes* informational content in those molecules. And the translation-realizing machinery of ribosomes and tRNA constitutes a consumer subsystem that *decodes* (and thereby exploits) that same informational content.

What we have in protein synthesis, then, is a producer-consumer economy of outcome-related, information-based transactions between homuncular subsystems. Such an arrangement surely warrants a representational interpretation, according to which the elements in which the homuncular subsystems deal are legitimately identified as coding for the outcomes in question.[11]

If we add these observations, about the architectural conditions under which a representational interpretation of some system is mandated, to our previous thinking about the purely causal conditions for representation, then the following general principle suggests itself: the presence of (i) systematic causal co-variation between the putative vehicles of content and specific causally downstream structures, (ii) arbitrariness, and (iii) systemic homuncularity is sufficient for coding-talk. (If one conceives of developmental systems in dynamical systems terms, then one might replace the causal co-variation condition with one that explicitly mentions developmental parameter-setting. If so, then the causal co-variation condition will be implicit, since elements that are rightly conceived as setting developmental parameters will always causally co-vary in a systematic way with the outcome states of interest.)

As it happens, my view is that conditions (i)-(iii) are not only jointly sufficient for representation, but necessary too. It seems undeniable that systematic causal co-variation is necessary for representation. The additional necessity of arbitrariness is, perhaps, clear enough. Thus, to give an intuitive non-genetic example, where the outcome in question is, say, keying my actions to the door-stopping potential of some book on my office shelf, the equivalence class of neural states which may perform the right outcome-achieving role of selecting a suitable book will be fixed precisely by the fact that some of those elements are able, when or-

[11] In [Wheeler, 2005, chapters 8 and 10] I argue that the interlocking architectural features of arbitrariness and homuncularity also form the basis of an adequate account of the notion of representation as used in cognitive science. In philosophy of mind and cognitive science, the connection between arbitrariness and representation has been made previously by, for example, [Pylyshyn, 1986], and the notion of homuncularity (or something very close to it) has been linked with representation before, by, for example, Millikan [1995]. The conceptual interlock between arbitrariness and homuncularity is not part of these theorists' treatments, although it is anticipated by Wheeler and Clark's [1999] link between arbitrariness and information-based consumption. The sense of homuncularity that I have pressed into service in this paper is superficially 'thinner' than its cognitive-scientific cousin (at least as I develop the latter), since the present notion does not *explicitly* require that the subsystems concerned be organized in an hierarchical manner. In fact however, in any homuncular analysis there will always be a background commitment to the idea that subsystems that perform relatively complex subtasks could, in principle, be analyzed into further subsystems that perform relatively simpler subtasks, until the whole edifice 'bottoms out' in subsystems that perform primitive bio-chemical functions. Thus there is always a (perhaps weak) sense of hierarchicality in play.

ganized and exploited in the right way, to carry some relevant item or body of information (e.g., that the book is heavy enough to hold the door open). Here it seems safe to say that the elements in question represent the associated worldly features. But now consider the outcome of simply holding my office door open. The equivalence class of suitable objects which may perform this role will be fixed by (roughly) the non-informational properties of being heavy enough and being sufficiently non-obstructive with respect to passing through the doorway. Here, where the equivalence class of different elements that could perform the function at issue is fixed by certain non-informational physical properties of those elements, there is simply no place for the language of representation. This suggests that arbitrariness is necessary for representation. And if, as my architecture-related reflections suggest, arbitrariness and homuncularity arrive on the explanatory scene arm in arm (conceptually speaking), then the claim that homuncularity is necessary for representation looks to be concurrently established.[12] So, if I am right, the joint presence of (i) systematic causal co-variation between the putative vehicles of content and specific causally downstream structures, (ii) arbitrariness, and (iii) systemic homuncularity is necessary and sufficient for coding-talk.

As my description of the machinery underlying protein synthesis indicates, that machinery satisfies conditions (i)–(iii). (Although I have not argued explicitly that there are appropriate causal co-variations in protein synthesis, it should be clear enough that there are systematic causal mappings between, on the one hand, both DNA and mRNA, and, on the other, proteins.) But how secure is the general account of representation that I have given? Here I shall consider four objections.

First, one might object to the claim that arbitrariness is necessary for representation, on the grounds that not all elements that we take to be representations have that property. This is the sort of complaint that needs to be settled on a case-by-case basis, but let's at least consider one of the more plausible candidates for positive representational status coupled with non-arbitrariness, namely onomatopoeiaic words.[13] Since the pronunciation of such words suggests their meaning (e.g., meow), it might seem that they cannot be arbitrary. Yet we still think of them as representational, so they provide a counter-example to my suggestion that arbitrariness is necessary for representation. However, it seems to me that the intuition that onomatopoeiaic words cannot be arbitrary trades on a thought that is not reflected in the concept of arbitrariness, as I have unpacked that concept here. Many different physical sound patterns could realize the word 'meow' (compare the way in which native French and native English speakers pronounce the word), and what fixes the equivalence class of appropriate sounds is the informational content that they carry (roughly, this is a sound that cats make). Of course, the class of sounds that may be legitimate physical realizers of the word 'meow' is presumably not infinite, but then infinite realizability is not required

[12] In [Wheeler, 1995, chapter 10] I give independent reasons for thinking that homuncularity is necessary for representation.

[13] This worry was put to me by Elliott Sober (in discussion). Thanks to Phyllis McKay and Peter Sullivan for helping me to think about the best way to repel it.

for arbitrariness, in the sense that I am using that term. The class of legitimate physical realizers of the so-called 'genetic code' is certainly not infinite.

Second, one might object to the claim that conditions (i)-(iii) are sufficient for coding, on the grounds that what is additionally necessary for (any sort of) representation is the presence of combinatorial structure — perhaps of a mild kind — in the inner elements, enabling structurally related elements to guide different-but-related outcomes. It is worth pointing out that the system underlying protein synthesis would plausibly satisfy this condition (for related thoughts, see [Godfrey Smith, 2000b]). However, as far I can see, and despite arguments to the contrary by, for example, Haugeland [1991], such systematicity concerns the *power of* a representational system, rather than its *status as* a representational system.

Third, one might complain that in moving beyond an austere causation-based story about coding, to one that is based on architectural features, I have introduced an ineliminable reference to function, and thus ultimately to natural selection. If so, then there would at least be a suspicion that I am open to the very criticisms of selection-based approaches that I myself have advanced. However although, in evolutionary biology, function-talk naturally invites an appeal to Darwinian selection, generating what we might call *Darwinian functions*, that is not the only way to think about functions in biological systems. *Causal role functions* [Cummins, 1975], as studied by, for example, anatomists and physiologists, are identified not by evolutionary history, but by analyzing an overall task (thinking, swimming, digesting food, assembling proteins) into well-defined subtasks performed by well-defined parts or subsystems. Griffiths illustrates the distinction with an example germane to our project here. A "sequence of nucleotides GAU has the [Darwinian function] of coding for aspartic acid if that sequence evolved by natural selection because it had the effect of inserting that amino acid into some polypeptide in ancestral organisms" [Griffiths, 2005, 1]. The same nucleotide sequence "has the [causal role function] of coding for aspartic acid if that sequence has the effect of inserting that amino acid into some polypeptide in the organism in which it occurs" [Griffiths, 2005, 2]. Homuncular analysis naturally buys into the causal role sense of function, but it remains a further issue whether or not the causal role function of an homuncular subsystem is accompanied by a function in the selective sense. But notice, in this context, that the notion of causal role function (which is conceptually richer that mere causal information) supports talk of misrepresentation, and thus plausibly of intentional information. Without additionally appealing to selection, we can surely make sense of a scenario in which intervening causes prevent the subsystemic outcome that is related to a particular causal role function from coming about.

Finally, one might worry that conditions (i)-(iii) suffer from their own excessive liberality problem, in that they will be met by inappropriate environmental factors. To see why this is plausibly not the case, we can build on an example due to Godfrey-Smith [2000a]. Take a plant that responds to an increase in day length by starting to flower. According to Godfrey-Smith, the connection between the cause (the increase in day length) and the effect (flowering) here is arbitrary, because

the cause could have been interpreted in many other ways by the flower. Thus the cause in this arrangement counts as arbitrary, it's environmentally located, and it looks like the kind of factor that really shouldn't count as a representation of a developmental outcome; so, with respect to the arbitrariness condition alone, excessive liberality emerges as a genuine danger. Of course, I have characterized arbitrariness not in terms of a cause potentially having a range of different effects, but in terms of the equivalence class of different physical factors that could have played the same causal role being fixed by informational rather than brutely physical considerations. Nevertheless, the content 'start to flower' could clearly have been carried by environmental factors other than increase in day length, so it looks as if arbitrariness in my sense is present too, and in the same worrying place. The solution (in the framework I am promoting) is to take seriously the conceptual interlock between arbitrariness and homuncularity with respect to the justification of coding-talk. For while, in the flower case, it might well be said that there is a consumer system that digests the putative information (by interpreting the increase in day length as an instruction to flower), it is hard to see how to make sense of the claim that the overall arrangement contains a producer system that has performed the role of *encoding that information in* the relevant causal factor, namely in the increase in day length. So the environmental factor in question does not emerge as being representational in character.

This response to Godfrey-Smith's example does not establish that non-genetic factors could not ever qualify as vehicles of representational content in development, once arbitrariness and homuncularity are plugged in as necessary conditions. Take animal signalling systems. If one could specify the appropriate causal co-variations (that is, between the signals and the construction of developmentally downstream structures), those systems will contain noises, marks, and so on, that will count as environmentally located vehicles of representational content. (Of course, the producer subsystem will be in one individual animal, while the consumer subsystem will be in another, but nothing I've said rules out such a state of affairs.) However, notice that the existence of such elements does not violate the weakened uniqueness constraint. It is neither unreasonable, nor extravagant, nor explanatorily inefficacious to claim that the developmental contribution of such factors is representational in character. What needs to be ruled out is the systematic inclusion of illegitimate factors (such as an increase in day length). And that, I think, is plausibly achieved by a proper recognition of the part played by the producer subsystem. However, that recognition also brings us to what, I suspect, is the most controversial claim that I shall make in this paper

5 A BULLET TO BITE

Strictly speaking, according to the proposal currently on the table, it's not the DNA molecules that constitute the representational vehicles that play a coding role in development, but rather the nucleotide triplets (the *codons*!) that make

up the mRNA molecules.[14] Genes don't code; mRNA does. Why is this the right unpacking of the proposal? On the view developed here, representation requires a producer and a consumer. The representations are the vehicles of content that support the communicative transactions between these systems. The producer encodes information into the vehicles in question, the consumer decodes information from them. In the case of protein synthesis, the consumer system is the distributed mechanism of ribosomes and tRNA that realises the process of translation in which mRNA determines the manufacture of proteins. So what is the producer system? The most compelling answer, it seems to me, is that it is the distributed mechanism underlying the process in which the organism's DNA acts as a template in the manufacture of *mRNA* molecules, that is, the producer system is the machinery of transcription and, in the case of eukaryotes, RNA splicing (as described above). It's that very machinery that encodes the information in mRNA molecules, the information that will later be decoded during translation. To see why this interpretation is the most compelling, we need to consider some objections.

The first is to claim that while there is a strict sense in which it's mRNA nucleotide triplets that code, the fact is that DNA codes *by extension*. Godfrey-Smith [2000b, 32] puts it like this: "The "genetic code" is, strictly speaking, the rule linking RNA base triplets with amino acids. This "interpretation" of the RNA determines the "interpretation" of the DNA from which the mRNA was derived." This suggestion faces a serious difficulty. To see why, we need to consider an analogy with the cognitive science of visually guided action. In the broadest terms, according to much thinking in cognitive science, patterns of stimuli on the retina determine the structure of certain inner states that intervene between sensing and action. Strictly speaking, what determines the final outcome (the agent's behaviour) will be some action-specifying inner state that needs ultimately to be translated into physical movements. Now if, as seems warranted, we map (a) the pattern of stimuli on the retina onto DNA sequences, (b) the process by which those stimuli determine the structure of the outcome-specifying inner state onto transcription plus RNA splicing, (c) the outcome-specifying inner states onto mRNA molecules, (d) the process by which those states are turned into physical behaviour onto translation, and (e) the behaviour onto proteins, then by something like the reasoning that Godfrey-Smith advocates in the case of protein synthesis, it would be right to say, in cognitive psychology, that patterns of stimuli on the retina code for particular actions. And that doesn't seem right. There will, of course, be systematic correlations between both (i) the form of the action and the retinal patterns, and (ii) the content of the inner action-specifying state and the retinal patterns, but the fact is that as we travel causally downstream from the

[14] Essentially the same claim is made, on related but importantly different grounds, by Bullock [1998]. Bullock treats genes as themselves encoders, a position which I reject (see later in this section), and he makes a pivotal appeal to natural selection in his argument that the machinery of protein synthesis contains a consumer system, an appeal which I think is unsustainable (see arguments in section 3 above).

retinal input, extra content is introduced that is relevant to the exact form of the actions produced. Crucially this content is introduced during the construction of the output-specifying inner states (reflecting, e.g., the goals and interests that the agent is pursuing, and that determine how the agent should respond to the input). Interestingly, in theories of visually-guided action where environmental stimuli are said to specify actions more directly (e.g., in Gibsonian ecological psychology), those theories are often characterised as being non-representational in character. The consequence of these observations is that one wouldn't have a mandate to say that how we interpret the inner action-specifying states here determines how we should interpret the retinal input. If the analogy holds, then similarly we should not endorse Godfrey-Smith's suggestion that the interpretation of the mRNA determines the interpretation of the DNA from which the mRNA was derived.

The obvious counter-move here is to question the analogy by claiming that nothing approaching the complexity present in the psychological case is present in the process by which DNA sequences are transformed into mRNA base triplets. Thus Maynard Smith [2000a] draws his own analogy, this time with Morse code. In the use of Morse code the content of the message is, Maynard Smith claims, first encoded into phonemes by the original coder (a human being), and then merely *converted into* Morse code. He then argues that, in the case of DNA, the original coder is natural selection, which encodes developmental information into genes. That information is then merely converted into mRNA base triplets.

The first thing to say here is that we have found good reasons to conclude that, in the present context, selection is not necessary for representation (see above), so the appeal to natural selection needs to be treated with suspicion. However, the claim about 'mere conversion' could in principle be freed from the link with natural selection. One might try to argue, for example, that the way in which the interpretation of the mRNA determines the interpretation of the DNA obviates need for a producer system altogether. What really needs to be resisted, then, is the claim that the DNA-to-mRNA transition can be relegated to anything approaching mere conversion on the phonemes-to-Morse model. The second point to make is that, in *the case of eukaryotes at least*, there are events that occur between transcription and the beginning of translation that undermine any such relegation. I have already mentioned RNA splicing. Sometimes this takes the form of so-called *alternative splicing* in which the same initial RNA transcript gets spliced in different ways to generate several proteins. In addition, there are other complex processes of RNA editing, involving the addition, removal, or replacement of bases. So, *in the case of eukaryotes at least*, the analogy with the mechanisms by which sensory stimulation results in inner action-coding seems to hold, which means that one cannot deploy Godfrey-Smith's strategy to establish that eukaryotic DNA codes in protein synthesis. And, having blocked the use of that strategy in the case of eukaryotes, it seems to me that we have good methodological reasons to extend our preferred interpretation — that mRNA not DNA codes in protein synthesis — to prokaryotes too. As we have seen, Godfrey-Smith himself concedes that *strictly speaking* the so-called genetic code is the

mapping between mRNA base triplets and amino acids, suggesting strongly that, *strictly speaking*, it's mRNA that codes for proteins. There seems little reason to speak loosely for prokaryotes if such talk has shown to be misleading in the case of eukaryotes.

What looks like an alternative way to resist my argument concerning the location of the coding entities in protein synthesis may be found in an argument due to Stegmann [2005]. Stegmann identifies a notion that he calls *instructional content*, unpacked as the information for the synthesis of some outcome, such that that outcome is determined via the step-by-step realization of operations specified in advance. The thought is that this kind of content is familiar from everyday representational entities such as cooking recipes and computer programs. Given this notion of content, Stegmann argues that if we look at the role of DNA in transcription, then we find that it carries instructional content, in virtue of the template-directed synthesis that produces (primary) RNA transcripts from DNA. Thus, if we take 'code for' to be equivalent to 'carries the information for,' genes get to code, *independently* of anything we might say about the relationship between DNA and proteins. The question then, is, can the coding relationship in transcription be extended *forwards*, so that it reaches proteins? Stegmann's answer is yes, but only under certain conditions. Here's the chain of thought: (a) just as DNA contains instructional content for synthesizing RNA transcripts, those transcripts contains instructional content for synthesizing proteins; (b) the bases in a DNA template stand in a neighbour relation to each other, in that C is next to T, T is next to G, and G is next to A; (c) the neighbour relation present in DNA is preserved in the RNA transcript, in that the base in the RNA product corresponding to C is next to the base in the RNA product corresponding to T, and so on; (d) this neighbour relation isn't disrupted in translation; so (e) the linear order of the DNA template determines the linear order of both the RNA and the protein; so (f) DNA codes for (carries the instructional content for) proteins.

Stegmann's argument, even if sound, is restricted in its scope. As we have seen, and as Stegmann himself notes, the no-disruption condition, (d), is typically not met for eukaryotes, so the putative result that genes code for proteins may well be restricted to organisms such as bacteria. Elsewhere the putative result is that genes code for RNA, RNA codes for proteins, but genes don't code for proteins. However, this is by-the-by, because there is a problem with Stegmann's argument. We think of cooking recipes and computer programs as having instructional content *only* because (i) a producer system — a cooking expert, a computer programmer — has encoded the instructional information in the physical vehicles which carry that information, and (ii) a consumer system — the cook using the recipe, the right compiler — interprets those physical vehicles as instructions. Notice that this is *not* a demand that there be sentient agents in the loop. As mentioned above, the systems that we rightly identify as producer systems and as consumer systems need not literally understand the content of the representations in question. They simply need to be play the right architectural roles. Nevertheless systems of this sort need to be part of the story. But if that's right, then DNA

doesn't carry instructional information, since (and this point has been bubbling just below the surface of my recent discussion), *there is no relevant producer system in the case of DNA*. Replication is not the same as encoding, so one cannot think of DNA as somehow *self*-encoding (with, of course, the help of some complex supporting chemical machinery). And if one tries to recruit natural selection as the producer system (cf. Maynard Smith's analogy with the Morse coder), one simply re-confronts the by-now familiar objection that factors which have not been selected for may sometimes qualify as coding within development. There would be no explanation for the positive representational status of such elements.

What this all suggests is that the part played by DNA in development is rather like the part played by sensory input in the perceptually guided action case. DNA doesn't code for outcomes, but rather provides a causally critical stimulus for subsequent development, a stimulus that is, of course, both determined by the target system's operational context (one which is environmental in the case of perceptual activity, and historical in the case of development), and partly predictive of the final outcome.

6 THE REACH OF THE CODE

The foregoing analysis of protein synthesis suggests that mRNA base triplets are rightly said to code for proteins. But do they also code for phenotypic traits? Some thinkers who have concluded that DNA codes for proteins have proceeded to worry that the reach of the code stops there, and that the claim that genes code for phenotypic traits is indefensible. Indeed, even prominent critics of the whole genetic coding bandwagon are often willing to grant that genes code for proteins, but not traits. Thus Griffiths claims that "the only truth reflected in the conventional view is that there is a genetic code by which the sequence of DNA bases in the coding regions of a gene corresponds to the sequence of amino acids in the primary structure of one or more proteins" [Griffiths, 2001, 395]. I shall bring the present treatment to a close then by considering an argument due to Godfrey-Smith (2000b) which questions the extension of the coding relationship from proteins to traits. If this argument is sound, it would compel me to conclude that mRNA codes only for proteins and not also for phenotypic traits. Here is the argument:

> The concept of genetic coding is now used to describe and distinguish the *entire causal paths* in which genes are involved. This use of the concept of genetic coding has, I claim, no empirical basis and makes no contribution to our understanding ...
>
> To make this claim is not to deny that at least some causal relations are transitive, and so to deny that genes can causally affect complex traits of whole organisms... The long causal reach of genes is not at issue in this paper. What is at issue is the relation of "coding for..." ... A case from everyday life illustrates the point. Suppose you know

> that if you order the extra-large pizza, that will have the consequence that the delivery arrives late. This fact does not imply that when you order the extra-large pizza you are also ordering them to make the delivery late. The likely or inevitable *effects* of a message are not all part of the *content* of the message. Similarly, genes can have a causal role which extends beyond the production of proteins, but proteins are all a gene can code for. [Godfrey-Smith, 2000b, 35]

Godfrey-Smith is surely right about at least one thing here. His pizza example does indeed show that the "likely or inevitable effects of a message are not all part of the content of the message". But, on the face of it, this doesn't provide a mandate for his conclusion that "proteins are all a gene can code for". It establishes only that we need a way of conceptually screening off those causally downstream (in this case, phenotypic) effects which *don't* count as part of the content of a particular coding from those that *do*. How might this be achieved, and the reach of the code thereby extended from proteins to traits? Notice that in providing an answer to *this* question, we are now at liberty to appeal to factors that we rejected when our target was a *different* question, namely 'Why should we use representational language *at all*, when trying to understand development? I have given an answer to this latter question, in terms of three conditions: appropriate causal co-variation, arbitrariness, and homuncularity. What I haven't done yet is give an answer to the former question, the question highlighted by Godfrey-Smith's argument. What Godfrey-Smith's pizza undoubtedly shows us is that the answer to that question, in the case of any outcome-directed representations, cannot be "Whatever the effects are that the representation in question has".

One initially attractive thought is that only the phenotypic outcomes that ensue in the normal developmental environment count as part of the content of the code. But that raises the thorny question of how 'normal' is to be interpreted here. It cannot be interpreted as 'statistically normal', since one of the lessons of Peter's inevitably late pizza is that some of the effects that an instruction has in its statistically normal environment may not be part of its content. Here it is tempting to revive an appeal to the intended effects, which would succeed in screening off the lateness of the pizza. (The intended effect was an on-time extra-large pizza, not a late one.) As we saw earlier, in the biological case, the appeal to intended effects will be unpacked in terms of selection. But now recall, once again, our (made-up) example of genetic hitchhiking, in which a non-selected-for gene that is causally implicated in the production of blue eyes hitchhikes into the population by being physically connected to a selected-for gene that is causally implicated in the production of a thick coat. Ignoring, for a moment, the matter of whether it's genes or mRNA nucleotide triplets that code, an appeal to selection will straightforwardly deliver the result that our representational element codes not only for proteins, but also for a white coat, since that trait (and thus the related coding element) has been selected for. But if we turn now to the coding element that is causally implicated in the presence of blue eyes, the appeal to selection leaves it devoid of any post-protein content, since blue eyes have not been selected

for. The discrepancy here indicates that the appeal to selection falls short of the explanatory mark, since it would surely be uncomfortable to be forced to conclude that the white-coat-related element codes for a white coat, whereas the blue-eyes-related element codes only for proteins. And thinking of the blue-eyes-related gene as being indirectly selected for certainly won't help, since the explanation for its presence is that having a thick coat is selectively advantageous in the environment in question; so that would make the content of the blue-eyes-related gene 'build a thick coat' which is surely not what we want.

Although there is undoubtedly more to be said on the reach of the code, the foregoing discussion indicates that it is a difficult and challenging issue. In view of the problems in extending that reach beyond proteins to traits, the default option ought to be to restrict it to proteins. Add this to the conclusion that the locus of coding talk in biological development is mRNA, the base triplets that determine the strings of amino acids constructed during protein synthesis, and the following picture emerges. The power of coding talk in development may be limited at both ends. Such talk doesn't stretch as far back as genes, and it may not stretch as far forward as phenotypic traits.

ACKNOWLEDGEMENTS

For important critical discussion of the ideas presented here, my thanks go to Andre Ariew, Andy Clark, Sharon Groves, Katie Kendig, Phyllis McKay, Matteo Mameli, Mohan Matthen, John Maynard Smith, Alan Millar, Marion Rosowski, Nicholas Shea, Eliott Sober, Ulrich Stegmann, Peter Sullivan, Terry Sullivan, Denis Walsh, and my 2005 *Foundations of Evolution* Masters class at the University of Edinburgh.

BIBLIOGRAPHY

[Bullock, 1998] S. Bullock. The emptiness of the self-contained coder: Commentary on an extract from 'Twisted Tales' by A. Clark. *Connexions*, 3: 7–9, 1998.

[Cummins, 1975] R. Cummins. Functional analysis. *Journal of Philosophy*, 72: 741–765, 1975.

[DeLisi, 1988] C. DeLisi. The human Genome Project. *American Scientist*, 76: 488–93, 1988.

[Dretske, 1981] F. Dretske. *Knowledge and the Flow of Information*. Cambridge MA: MIT Press, 1981.

[Dupre, 2005] J. Dupre. Are there genes? In A. O'Hear (ed.), *Philosophy, Biology and Life*. Supplement to *Philosophy*, 56. Cambridge: Cambridge University Press, pages 193–210, 2005.

[Godfrey-Smith, 2000a] P. Godfrey-Smith. Information, arbitrariness, and selection: Comments on Maynard Smith. *Philosophy of Science*, 67: 202–207, 2000.

[Godfrey-Smith, 2000b] P. Godfrey-Smith. On the theoretical role of genetic coding. *Philosophy of Science*, 67: 26–44, 2000.

[Goodwin, 1994] B. Goodwin. *How the Leopard Changed its Spots: the Evolution of Complexity*. London: Phoenix, 1994.

[Griffiths, 2001] P. E. Griffiths. Genetic information: A metaphor in search of a theory. *Philosophy of Science*, 68: 394–412, 2001.

[Griffiths, 2005] P. E. Griffiths. Function, homology and character individuation. Paper presented at 'The concept of function in biology and language', Universitá di Catania, May 2005.

Available at:
http://philsci-archive.pitt.edu/archive/00002351/01/Func.Hom.Chr.PDF

[Griffiths and Gray, 1994] P. E. Griffiths and R. D. Gray. Developmental systems and evolutionary explanation. *Journal of Philosophy*, 91(6): 277–304, 1994.

[Griffiths and Knight, 1998] P. E. Griffiths and R. D. Knight. What is the developmentalist challenge? *Philosophy of Science*, 65: 253–258, 1998.

[Griffiths and Stotz, forthcoming] P. E. Griffiths and K. Stotz. Gene. In M. Ruse and D. Hull (eds.), *Cambridge Companion to Philosophy of Biology*. Cambridge: Cambridge University Press, forthcoming.

[Halder *et al.*, 1995] G. P. Halder, P. Callaerts and W. J. Gehring. Induction of ectopic eyes by targeted expression of the *Eyeless* gene in *Drosophila*. *Science*, 267: 1788–1791, 1995.

[Haugeland, 1991] J. Haugeland. Representational genera. In W. Ramsey, S. P. Stich and D. E. Rumelhart (eds.), *Philosophy and Connectionist Theory*. Hillsdale, New Jersey: Lawrence Erlbaum, pages 61–90, 1991.

[Lorenz, 1965] K. Lorenz. *Evolution and the Modification of Behaviour*. Chicago: University of Chicago Press, 1965.

[Mahner and Bunge, 1997] M. Mahner and M. Bunge. *Foundations of Biophilosophy*. Berlin, Heidelberg, New York: Springer-Verlag, 1997.

[Mameli, 2004] M. Mameli. Nongenetic selection and nongenetic inheritance. *British Journal for the Philosophy of Science*, 55(1): 35–71, 2004.

[Mameli, 2005] M. Mameli. The inheritance of features. *Biology and Philosophy*, 20(2-3): 365–399, 2005.

[Maturana and Varela, 1987] H. Maturana and F. J. Varela. *The Tree of Knowledge: the Biological Roots of Human Understanding*. Boston: New Science Library, 1987.

[Maynard Smith, 1998] J. Maynard Smith. *Shaping Life: Genes, Embryos and Evolution*. London: Weidenfeld and Nicolson, 1998.

[Maynard Smith, 2000a] J. Maynard Smith. The concept of information in biology. *Philosophy of Science*, 67: 177–194, 2000.

[Maynard Smith, 2000b] J. Maynard Smith. Reply to commentaries. *Philosophy of Science*, 67: 214–218, 2000.

[Millikan, 1995] R. G. Millikan. *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: MIT Press, 1995.

[Odling-Smee *et al.*, 2003] F. J. Odling-Smee, K. N. Laland and M. W. Feldman. *Niche Construction; the Neglected Process in Evolution*, Princeton: Princeton University Press, 2003.

[Oyama, 1985] S. Oyama. *The Ontogeny of Information*. Cambridge: Cambridge University Press, 1985.

[Oyama *et al.*, 2001] S. Oyama, P. E. Griffiths, and R. D. Gray. *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge, MA.: MIT Press, 2001.

[Pylyshyn, 1986] Z. W. Pylyshyn. *Computation and Cognition*. Cambridge, Mass.: MIT Press, 1986.

[Sarkar, 2000] S. Sarkar. Information in genetics and developmental biology: Comments on Maynard Smith. *Philosophy of Science*, 67: 208–213, 2000.

[Sarkar, 2005] S. Sarkar. How genes encode information for phenotypic traits. In his *Molecular Models of Life: Philosophical Papers on Molecular Biology*. Cambridge MA.: MIT Press, pages 261–283, 2005. This is a revised and expanded version of a paper that appears in C. Hitchcock (ed.), *Current Issues in the Philosophy of Science*, Oxford: Blackwell, pages 259–274, 20004.

[Shannon and Weaver, 1949] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[Stegmann, 2005] U. E. Stegmann. Genetic information as instructional content. *Philosophy of Science*, 72: 425–443, 2005.

[Sterelny, 1995] K. Sterelny. Understanding life: Recent work in philosophy of biology. *British Journal for the Philosophy of Science*, 46: 155–183, 1995.

[Sterelny, 2000] K. Sterelny. The "Genetic Program" program: A commentary on Maynard Smith on information in biology. *Philosophy of Science*, 67: 195–201, 2000.

[Sterelny and Griffiths, 1999] K. Sterelny and P. E. Griffiths. *Sex and Death: an Introduction to Philosophy of Biology*. Chicago: University of Chicago Press, 1999.

[Sterelny and Kitcher, 1988] K. Sterelny and P. Kitcher. The return of the gene. *Journal of Philosophy*, 85(7): 339–361, 1988.

[Wheeler, 2003] M. Wheeler. Do genes code for traits? In A. Rojszczak, J. Cachro and G. Kurczewski (eds.), *Philosophical Dimensions of Logic and Science: Selected Contributed Papers from the 11th International Congress of Logic, Methodology, and Philosophy of Science. Synthese Library* vol. 320. Kluwer, pages 151–164, 2003.

[Wheeler, 2005] M. Wheeler. *Reconstructing the Cognitive World: the Next Step*. Cambridge MA: MIT Press, 2005.

[Wheeler and Clark, 1999] M. Wheeler and A. Clark. 'Genic representation: Reconciling content and causal complexity. *British Journal for the Philosophy of Science*, 50(1): 103–135, 1999.

[Wood *et al.*, 1976] D. J. Wood, J. Bruner, and G. Ross. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17: 89–100, 1976.

This page intentionally left blank

Part IV
Taxonomy

This page intentionally left blank

SPECIES, TAXONOMY, AND SYSTEMATICS

Marc Ereshefsky

1 INTRODUCTION

Conceptual issues in biological taxonomy often straddle the boundary between biology and philosophy. Consider questions over the nature of species. Are species natural kinds or individuals? Does 'species' refer to a category in nature or is it merely a useful device for organizing our biological knowledge? These questions involve elements of metaphysics, epistemology, and philosophy of science. These questions also turn on empirical information and biological theory. When we ask about the nature of species are we asking a philosophical or biological question? The answer is 'both.'

Three conceptual issues in biological taxonomy and systematics will be discussed in this chapter. One is the ontological status of species. Many if not most philosophers believe that species are natural kinds –classes of organisms with theoretically significant similarities. Other philosophers, and many biologists, believe that species are individuals akin to particular entities. The outcome of this debate has implications beyond biological taxonomy, for example, it has implications for our conception of human nature. Is *Homo sapiens* a kind defined by similarities among its members, or is *Homo sapiens* an evolving lineage defined by genealogy? If species are individuals, then no qualitative trait is necessary and sufficient for being a human.

Another conceptual issue is taxonomic pluralism. A common scientific and philosophical view is that, from a God's eye perspective, there is a single correct classification of the organic world. A number of authors respond that this view is wrong. They argue that there is no single correct classification of the organic world and that the organic world itself is pluralistic. This issue has implications outside of taxonomy. In philosophy of science it raises the question of whether philosophers should promote scientific unity as an aim of science. If taxonomic pluralism occurs in biology then biology, and science as a whole, lacks unity.

A third conceptual issue concerns the Linnaean hierarchy: should biologists keep using the Linnaean hierarchy or should they adopt a new system of classification? The Linnaean hierarchy was developed prior to the Darwinian revolution. Many biologists and some philosophers believe that the Linnaean hierarchy faces pressing problems. Some even suggest junking the Linnaean hierarchy. Again, important issues outside of classification hinge on this debate. Most biodiversity studies are given in terms of Linnaean ranks, for example, the rank of family. If those ranks do

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

not correspond to categories in nature, as many critics of the Linnaean hierarchy argue, then most biodiversity studies are based on faulty measures.

1.1 Terms and Distinctions

We have already used the terms *classification*, *taxonomy*, and *systematics*, but what do they mean? What distinctions do they highlight? We are familiar with classifications: organisms are sorted into species, species are sorted into genera, genera into families, and so on up the Linnaean hierarchy. A classification is a hypothesis concerning how organisms or taxa are related. Taxonomy is the discipline that tells us how to sort organisms into taxa and taxa into more inclusive taxa. Taxonomic theory offers principles for constructing classifications. Systematics does not tell us how to construct classifications but studies how organisms and biological taxa are related in world. Classification, taxonomy, and systematics line up in the following way. Systematics, the study of the relations among organisms and taxa, guides our choice of taxonomic theory; taxonomic theory guides biologists in constructing classifications of the organic world. The path from systematics to classification, however, is far from smooth. Biologists do not agree on systematic theories, so they posit different taxonomic theories (i.e., different principles for constructing classifications). Moreover, even when there is agreement on taxonomic theory, how to implement that theory can be controversial. For example, biologists tend to prefer more parsimonious classifications, but they disagree on how to measure parsimony.

Another distinction is between *species taxa* and the *species category*. Species taxa are groups of organisms; *Homo sapiens* and *Drosophila melanogaster* (a fruit fly) are examples of species taxa. The species category is a more inclusive entity: it is the class of all species taxa. When biologists and philosophers discuss the definition of 'species' they are discussing the definition of the species category. In Section 2, we focus on the ontological status of species *taxa*: is a species taxon a kind or an individual? In Section 3, we focus on the species *category*: is there a single, unified species category? In Section 4 we move up the Linnaean hierarchy and discuss the reality of the other Linnaean categories –genera, families, and so on. Does the Linnaean hierarchy reflect a hierarchy of natural categories, or are the Linnaean ranks merely useful instruments for organizing information about the organic world?

2 SPECIES

2.1 Essentialism

Traditionally philosophers have treated biological species as natural kinds with essences. This approach to species is found in the work of Kripke [1972] and Putnam [1975], and has its roots in Aristotle and Locke. Biologists have been essentialists as well. John Ray, Maupertuis, Bonnet, Linnaeus, Buffon, and Lamark

all talk of classifying species taxa according to their species-specific essences [Hull, 1965; Sober, 1980; Mayr, 1982].[1] The view that species are kinds with essences is part of a general view — kind essentialism. A natural kind is a class of entities that share a kind-specific essence. Such essences capture the fundamental structure of the world; or using Plato's phrase, they "carve nature at its joints."

Kind essentialism has two crucial tenets. First, all and only the members of kind share a common essential property. If an entity has a certain essence, it is member of a particular kind. If, for example, a piece of metal has a certain atomic structure, it is a member of the kind gold. However, some properties occur in all and only the members of kind yet they are not that kind's essence. The Aristotelian example is humans and the property featherless biped. All and only humans are featherless bipeds, nevertheless that property is not the essence of humans. Here is where the second tenet of essentialism comes into play. The essence of a kind is a property that plays a fundamental role in explaining the occurrence of other properties typically found among the members of a kind. For example, the atomic structure of gold causes pieces of gold to dissolve in acid and conduct electricity. That atomic structure, not any property that occurs in all and only pieces of gold, is the essence of the kind gold.

Both philosophers and biologists have applied kind essentialism to species. There are few essentialist biologists today, but essentialism remains a common view among philosophers. A number of philosophers and biologists have challenged species essentialism — the view that species are natural kinds with essences [Mayr, 1959; Hull, 1965; Ghiselin, 1974; Sober, 1980; Dupré, 1981]. Here are two lines of argument against species essentialism.

Recall the first tenet of essentialism: a property must occur in all and only the members of a kind. Biologists have been hard-pressed to find a biological trait that occurs in all and only the members of a particular species. Evolutionary theory explains why. Suppose a genetically based trait were found in all the members of a species. Mutation, random drift, and recombination can cause that trait not to occur in a future member of that species. All it takes is the disappearance of a trait in a single member of a species to show that it is not essential. The universality of a trait among the members of a species is quite fragile.

Evolutionary forces also undermine the uniqueness of a trait among the members of a species. Different species frequently live in similar habitats that cause the parallel evolution of similar traits in different species. Birds and bats, for example, each have wings, but the evolutionary path for each type of wing is distinct. Organisms in different species also share common ancestors, so they draw

[1] Windsor [2003] has challenged the claim that prominent pre-Darwinian biologists were essentialists. This raises an interesting issue concerning taxonomic thought. Linnaeus, for example, wrote that organisms should be classified according to their taxon-specific essences. However, he realized we do not have access to those properties, so he classified according to clusters of similarities [Ereshefsky, 2001]. Was Linnaeus an essentialist or not? Linnaeus advocated essentialism but could not implement it in practice. (Biologists often advocate taxonomic theory that they cannot yet use.) So perhaps the proper answer is, in some ways Linneaus was an essentialist, in other ways he was not.

on common stores of genetic material and developmental constraints. These common genetic and developmental resources cause the members of different species to share similar characteristics. Just think of all those organisms in the world with four limbs. Just as the forces of evolution work against the universality of a trait in a species they also work against the uniqueness of trait in a species.

The above considerations pose a strong challenge to the essentialist requirement that a trait must occur in all and only the members of a species. This does not mean that the occurrence of a biological trait in all and only the members of species is empirically impossible. Nonetheless, consider what conditions must be met for that requirement to hold. For essentialism to work, a trait must occur in all the members of species, for the entire life of that species — from its initial speciation event to its extinction. Moreover, for a trait to qualify as an essential property of a species, it must occur only in that species. That is, the trait in question cannot occur in any member of any other species for the entire span of life on this planet, indeed, for the entire span of life in this universe. This is a tall order! The occurrence of a trait in all and only the members of species is an empirical possibility. But given current biological theory, that possibility is unlikely.

Sober [1980], following Mayr [1969], offers a different objection to species essentialism. Sober illustrates how essentialist explanations have been replaced by evolutionary ones in biology, thereby rendering essentialist explanations theoretically obsolete. Suppose we want to explain variation in a population, say variation in height. The essentialist explanation cites the essence of the organisms in a species, and then cites ontogenetic interference that prevents the manifestation of that essence in some or all members. Organisms of that species would have the same height if they were exposed to the same environment, but they are exposed to different environments, hence there is variation from a common type. Variation is explained by essence and interference. In contrast, the evolutionist explains variation in a population without positing essences. Evolutionists cite the gene frequencies within a population and the evolutionary forces that affect what frequencies occur in the next generation. Variation in height is explained by citing the relevant gene frequencies in one generation plus the occurrence of such evolutionary forces as selection and drift. No essences are posited. In contemporary biology, the positing of species essences has become theoretically superfluous.

2.2 Species as Individuals

If species taxa are not natural kinds with essences, what is their ontological status? The prevailing view is that species are not natural kinds but individuals. Ghiselin [1974] and Hull [1978], the most prominent advocates of this view, contrast natural kinds and individuals in terms of spatiotemporal restrictedness. Membership in a kind requires that the members of a kind share a set of qualitative properties. A drop of liquid is water so long as it has the molecular structure H_2O. It does not matter where that liquid is located, whether it is on Earth now or in a distant galaxy in a million years. So long as that drop has the molecular structure H_2O

it is water. Individuals, unlike kinds, consist of parts that are spatiotemporally restricted. Consider a paradigmatic individual, a mammalian organism. The parts of that organism cannot be scattered around the universe at diverse times if they are parts of a single living organism. Various biological processes, such as respiration and digestion, require those parts to be causally and spatiotemporally connected. The parts of a mammal must exist in a particular space-time region. Generalizing from these examples, the parts of an individual are spatiotemporally restricted, whereas the members of a kind are spatiotemporally unrestricted.

According to Ghiselin and Hull, membership in a species is spatiotemporally restricted, thus species are individuals and not natural kinds. What is their argument that species are spatiotemporally restricted entities? The argument starts with the assumption that 'species' is theoretical term in evolutionary theory, so the ontological status of species is determined by the role 'species' plays in evolutionary theory. According to Hull [1978], species are units of evolution in evolutionary biology, and as such they are individuals. A number of processes can cause a species to evolve, natural selection is just one such process. Minimally, selection can cause a species to evolve by changing its gene frequencies from one generation to the next. More significantly, selection will cause a species to evolve by causing a rare trait to become prominent after a number of generations. Suppose the second type of evolution occurs. For such evolution to occur, the selected trait must be passed down through the generations of a species. Traits are not inherited unless some causal connection exists between the members of a species. Sex and reproduction require that organisms, or their parts (gametes, DNA), come into contact. Thus, evolution by selection requires the generations of a species to be causally and spatiotemporally connected. The organisms of a species cannot be scattered throughout the universe, but must occupy a spatiotemporally restricted region. Given that species are units of evolution, they are individuals and not kinds.

The species are individuals thesis is the prevalent ontological view of species. Nevertheless, that view has been contested on a number of grounds. Ruse [1987] and Ereshefsky [1991] argue that many species are not individuals because they fail to have the requisite cohesion required of individuals. Many species consist of asexual organisms. Such organisms are connected through parent-offspring relations, but no exchange of genetic material exists among the members of a species in a single generation. Consequently, asexual species are merely historical entities — their members are connected to a common ancestor, but there is no ongoing causal connection among contemporaneous members. This is certainly right, but suggesting that it shows that species are not individuals misses the importance of the individuality thesis. The important distinction is between being a natural kind and being an individual. Even as mere historical entities species are spatiotemporally restricted entities and not kinds. The idea that species are defined by causal connections rather than qualitative similarity is the important claim of the species are individuals thesis.

Another criticism of the individuality thesis challenges the assertion that all species are spatiotemporally restricted entities. Kitcher [1984], for example, be-

lieves that some species taxa are spatiotemporally continuous entities, while other species taxa are spatiotemporally scattered groups whose members share structural similarities. Such species are defined by their members having theoretically important genetic, chromosomal, and developmental similarities. Thus, Kitcher believes that there are two types of species: "historical" species and "structural" species. Historical species are defined by genealogy; structural species are defined by significant similarity. A problem with Kitcher's account of species is that it is out of step with contemporary biological practice. Biologists do not posit structural species. A quick glance at a biology text reveals that historical approaches to species, not structural approaches, are the going concern in biology. The question of whether species are kinds or individuals is a question about how the theoretical term 'species' functions in biology. Kitcher's argument against all species being individuals posits an account of species that is outside of contemporary biology.

2.3 Homeostatic Property Cluster Kinds

Another response to the species are individuals thesis is offered by proponents of an alternative approach to natural kinds. According to Boyd [1999a; 1999b], Griffiths [1999], Wilson [1999], and Millikan [1999], species are natural kinds on a proper conception of natural kinds. These authors adopt Boyd's Homeostatic Property Cluster (HPC) theory of natural kinds. HPC theory assumes that natural kinds are groups of entities that share stable similarities. HPC theory does not require that species are defined by traditional essential properties. The members of *Canis familaris*, for example, tend to share a number of common properties — having four legs, two eyes, and so on, but given the forces of evolution, no biological property is essential for membership in that species. For HPC theory, the similarities among the members of a kind must be stable enough to allow better than chance prediction about various properties of a kind. Given that we know that Sparky is a dog, we can predict with greater than chance probability that Sparky will have four legs.

HPC kinds are more than groups of entities that share stable clusters of similarities. HPC kinds also contain "homeostatic causal mechanisms" that are responsible for the similarities found among the members of a kind. The members of a biological species interbreed, share common developmental programs, and are exposed to common selection regimes. These "homeostatic mechanisms" cause the members of a species to have similar features. Dogs, for instance, tend to have four legs and two eyes because they share genetic material and are exposed to common environmental pressures. An HPC kind consists of entities that share similarities induced by that kind's homeostatic mechanisms. According to Boyd, species are HPC kinds and thus natural kinds because "species are defined . . . by . . . shared properties and by the mechanisms (including both "external" mechanisms and genetic transmission) which sustain their homeostasis" [1999b, 81].

HPC theory provides a more promising account of species as natural kinds than essentialism. HPC kinds need not have a common essential property, so

the criticisms of species essentialism are avoided. Furthermore, HPC allows that external relations play a significant role in inducing similarity among the members of a kind. Traditional essentialism assumes that the essence of a kind is an internal or intrinsic property of a kind's members, such as the atomic structure of gold or the DNA of tigers. Such intrinsic essences are ultimately responsible for a kind's similarities. HPC theory is more inclusive because it recognizes that both the internal properties of organisms and the external relations of organisms are important causes of species-wide similarities. For instance, HPC theory but not essentialism cites interbreeding as a fundamental cause of similarity among the organisms of many species.

While HPC theory is better at capturing the features of species than essentialism, a question remains: Does HPC theory provide an adequate account of species as natural kinds? Some argue that it does not [Ereshefsky and Matthen, 2005]. Here are two problems with HPC theory. HPC theory's objective is to explain the existence of stable similarities within groups of entities. However, species are also characterized by persistent differences. While HPC theory gives an account of the similarities among the members of a species, it provides no account of the persistent differences among a species' members. As such, it provides an impoverished account of species. Polymorphism — stable variation within a species — is an important feature of nearly every species. Species polymorphisms are easy to find. Consider sexual dimorphism. Within any mammalian species there are pronounced differences between males and females. Or consider polymorphism in the life cycles of organisms. The lives of organisms consist of dramatically different life stages, such as the difference between the caterpillar and butterfly stages of a single organism. HPC theorists recognize the existence of polymorphism, but they do not recognize polymorphism as a central feature of species in need of explanation. HPC theorists only privilege and attempt to explain similarities. In addition to Boyd's 'homeostatic' mechanisms we need to recognize 'heterostatic' mechanisms that maintain species variation.

A second problem with an HPC account of species concerns the identity conditions of species. The members of a species vary in their traits. Moreover, they vary in their homeostatic mechanisms. Over time and across geographic regions, the members of a single taxon are often exposed to different homeostatic mechanisms [Ereshefsky and Matthen, 2005]. Given such variation, what causes organisms with different traits and exposed to different homeostatic mechanisms to be members of the same species? The common answer is genealogy: the members of a species taxon form a continuous genealogical entity on the tree of life. A species' homeostatic mechanisms are mechanisms of one species because they affect organisms that form a unique lineage.

Boyd and promoters of HPC theory recognize that the importance of genealogy and see historical relations as one type of homeostatic mechanism. However, Boyd does not see genealogy as the defining aspect of species taxa, and this goes against a fundamental assumption of biological systematics: that species are first and foremost continuous genealogical entities. Consider Boyd's example of a species

that arose through multiple hybridization events. Boyd [1999b , 80] characterizes this species as containing "distinct lineages" of organisms that have significant similarities. This species is not a single historical entity, but a collection of separate lineages with separate origins. Boyd is quite clear that similarity and not genealogical connectedness is the final arbitrator of species sameness. This assumption makes sense given that Boyd believes that species are kinds and kinds are ultimately similarity-based classes that play a role in induction. But this view of the identity conditions of species taxa conflicts with the standard view in biological systematics that species are continuous genealogical lineages. Boyd, like other participants in the debate over the ontological status of species, assumes that 'species' is a theoretical term in evolutionary biology, so the status of species is determined by their role in biological theory. Nevertheless, HPC theory's preference for similarity-based kinds conflicts with the biological assumption that species are genealogical entities. Consequently, HPC theory fails to capture the proper identity conditions for species taxa.

2.4 Population Structure Theory

An alternative approach to species is offered by Ereshefsky and Matthen's [2005] "Population Structure Theory" (PST). PST treats similarity as just one type of trait distribution in species. PST does not privilege similarity over polymorphism, so PST offers a more inclusive account of trait distributions in species than HPC theory. In addition, PST highlights a common type of explanation in biology — one that cites the population and inter-population structures of species. Such *population structure explanations* explain trait distributions in species, whether those distributions involve similarity or dissimilarity.

Population structure explanations are pervasive in biology. Consider E. O. Wilson's [1968] explanation of different social castes in some insect species. The fitness of an insect colony is enhanced by its ability to respond to certain sorts of contingency. There are several types of such contingency, and specialization is required to deal with each of them, so there is selection for different castes. In addition, there is selection for an optimal mix of castes. The optimal mix of castes is calculated by figuring out how to keep the combined cost of various contingencies at a tolerable level. Wilson's theory explains the distribution of difference and why some castes occur in small numbers and others occur in large numbers. Wilson attributes the distribution of castes to group selection: different colonies with different phenotypic distributions compete against one another, and the ones with more optimal distributions are selected. Wilson's account explains population variation not uniformity. Furthermore, it explains such variation by citing the structures of populations — here the distributions of castes in a colony.

Wilson's explanation depends on group selection, but explanations citing population structure need not appeal to group selection. Consider a population structure explanation that does not involve group selection and is aimed at explaining similarity within a species. A prime example is sexual dimorphism. Male elk have

a number of similarities, including having large, fuzzy antlers. What explains that similarity? One cause — the proximal cause — is the individual development of each male elk. A deeper explanation — the distal cause — turns on relationships between male and female elks. Male antlers are the result of sexual selection. Such selection requires the participation of both male and female elk. Looked at in this way, we see that the existence of similarities within lower level groups, here within the genders, depend on higher level groups (here species) and the diversity within them. That is, polymorphism at the higher level, and the population structure that binds polymorphism, is essential in explaining lower level similarities within the genders and other sub-groups of a species.

Population structure explanations are common, and arguably essential, for understanding diversity and similarity within species. Such explanations are also essential for understanding the identity conditions of species. As we have seen, species are first and foremost genealogical entities. Genealogy is an inter-population structure — species are lineages of populations. So according to biological systematics, species identity is defined in terms of population and inter-populational structures, not organismic similarity. PST theory, unlike HPC theory, properly captures the identity conditions of species. Stepping back, we see that PST has three virtues. First, it accounts for both similarity and polymorphism within species. Second, by citing population and inter-population structures, PST provides a fundamental explanatory schema for explaining trait distributions in species. Third, that explanatory schema provides the proper basis for understanding species identity.

3 TAXONOMIC PLURALISM

So far, we have discussed the nature of species taxa. Let us now move up one level and ask about the nature of the species category. Typically, biologists and philosophers believe that there is one correct definition of that category; they believe that there is one correct 'species concept,' as biologists call it. However, the biological literature contains over a dozen species concepts [Claridge, Dawah, and Wilson, 1997]. These concepts are not fringe or crank concepts, but concepts proposed and investigated by prominent biologists. Biologists and philosophers have taken two stances to this plethora of species concepts. Monists believe that one concept is the correct species concept, or one concept is to be preferred over all others. Monists believe that biologists need to sift through the various proposed species concepts and determine which concept gives the proper definition of 'species.' Pluralists take a different stance. They believe that more than one species concept is worthy of acceptance. This section explores the arguments for and against species pluralism.

3.1 The Case for Pluralism

Let us start by introducing three prominent species concepts in biology. There are many more prominent species concepts, but introducing three is sufficient for providing the argument for pluralism. The most common species concept in the biological literature is Mayr's [1970] Biological Species Concept. The Biological Species Concept defines a species taxon as a group of organisms that can successfully interbreed and produce fertile offspring. According to that concept, a species' integrity is maintained by interbreeding within a species as well as by reproductive barriers between organisms in different species. The Ecological Species Concept defines a species taxon as a lineage of organisms maintained and segmented by ecological forces [Van Valen, 1976]. Stabilizing selection maintains a species' integrity, while disruptive selection can lead to new species. The Phylogenetic Species Concept (which has multiple versions) defines a species taxon as a basal monophyletic lineage [Mishler and Brandon, 1987]. A monophyletic lineage contains all and only the descendants of a common ancestor. Because monophyletic lineages occur up and down the Linnaean hierarchy, species are defined as *basal* monophyletic lineages — the smallest lineages represented in Linnaean classifications.

These species concepts, the biological, ecological, and phylogenetic, not only provide different definitions of 'species,' their use gives rise to different classifications of the organic world. This is confirmed by numerous empirical investigations. The most glaring discrepancy is between the Biological Species Concept (BSC) and the other two concepts. BSC requires that the organisms of a species exchange genetic information through interbreeding. That requires sexual reproduction. BSC does not require that every member of a species successfully interbreed, but it does require that a sufficient number of the organisms sexuality reproduce to maintain a species' integrity. The problem is that most of life on this planet does not reproduce sexually but asexually, through cloning or vegetative means. Asexual organisms do not form species according to the BSC. Nevertheless, asexual organisms do form species according to the Phylogenetic Species Concept (PSC) and Ecological Species Concept (ESC). For the PSC, species are defined genealogically, independent of mode of reproduction. For the ESC, species are defined as lineages of organisms maintained by selection forces. PSC's and ESC's classifications of the organic world include asexuals, while BSC's classifications exclude asexuals. These species concepts carve up the world in different ways.

Other cases of species pluralism are more complicated. For example, the BSC and the PSC sort the very same organisms into different species. Consider the case of ancestral species. Many supporters of the BSC believe that a standard form of speciation occurs when a population of a species becomes isolated from the main body of a species and undergoes a 'genetic revolution.' The parental species, or 'ancestral species,' remains intact. For proponents of the BSC, two species are present in such cases: the ancestral species consisting of A and B, and the new species C, see Figure 3.1.

However, the PSC cannot allow the existence of two species in this case. Recall

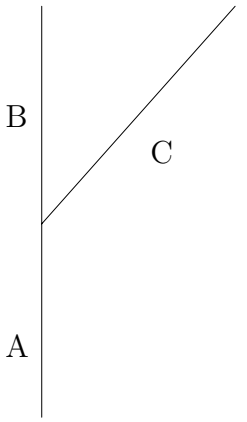


Figure 1. A branching event on a phylogenetic tree. If species must be monophyletic, then A+B cannot form a species. Some of the descendant's of A+B's ancestor are not contained in A+B but are in C.

for the PSC, a species must contain all and only the descendents of a common ancestor. The ancestral species consisting of A and B violates that requirement on species taxa: some of its descendents belong to the new species C. Thus, according to the PSC, either there is one species present (the combination of A, B, and C), or there are three species (the ancestral species A, which went extinct, and two new species, B and C). Either way, the PSC and the BSC cross-classify the very same group of organisms. Take an organism, X, in B. According to the BSC, X belongs to a species consisting of A and B. According to the PSC, X either belongs to a species containing only B or a species containing A, B, and C. Each species concept places X into two different taxa.

The above examples are just the tip of the iceberg of examples where species concepts provide different classifications of the same group of organisms. Generalizing from these examples, different species concepts give rise to different classifications of the organic world. Pluralists believe that examples like these show that we should take a pluralistic approach to biological classification: different species concepts provide different but equally legitimate classifications. Monists disagree. Before turning to monist responses, let us focus on the various brands of pluralism in the literature. This will help further articulate the pluralist's argument.

3.2 Varieties of Pluralism

The argument for pluralism suggested above is motivated by ontological and not epistemological considerations. Some authors (for example, Rosenberg [1994]) suggest that we adopt pluralism because of our epistemological limitations. The world is exceedingly complex and we have limited cognitive abilities, so we should accept a plurality of simplified classifications of the world. The taxonomic pluralism

advocated by Kitcher [1984], Ereshefsky [1992], and Dupré [1993] is not epistemologically driven. It is motivated by the idea that evolutionary theory, a well substantiated theory, tells us that the organic world is multifaceted. Pluralism is an ontological implication of one of our best scientific theories.

Though Kitcher, Ereshefsky, and Dupré agree that species pluralism should be adopted for ontological reasons, they adopt different versions of pluralism. Ereshefsky's form of pluralism, as outlined in the previous section, asserts that the tree of life is divided into different types of lineages — interbreeding species, ecological species, and phylogenetic species. Furthermore, these lineages cross-classify organisms on the tree of life (as illustrated by the examples in the previous section). Why are there different types of species lineages? They are the result of different evolutionary forces: interbreeding species are the result of interbreeding; ecological species are caused by natural selection; and, phylogenetic species are the result of genealogy. To highlight only one or two of these types of lineages is to give an impoverished account of evolution. Of course, this picture of evolution could be mistaken, but it is our current best picture of evolution.

Kitcher and Dupré adopt a different version of ontological pluralism than the pluralism advocated by Ereshefsky. Both Kitcher and Dupré recognize the three species concepts explored here — the biological, ecological, and phylogenetic species concepts. Kitcher and Dupré also accept species concepts based on shared similarities among the members of a species even when such species do not form continuous genealogical entities. For example, Kitcher posits 'structural' species concepts that allow species to be spatiotemporally disconnected entities (see Section 2.2). As should be familiar by now, 'species' is a theoretical term in biology, thus the ontological status of species taxa is determined by the role that term plays in biological theory. According to evolutionary biology, species taxa are first and foremost genealogical entities. A quick search of the biological literature shows that the species concepts suggested by biologists assume that species taxa form continuous genealogical entities. Given these considerations, the form of species pluralism advocated by Kitcher and Dupré is too liberal. It posits species concepts outside of contemporary biology.

Another type of species pluralism is suggested by Mishler and Brandon [1987]. While Kitcher and Dupré's form of pluralism is too liberal, Mishler and Brandon's is too conservative. Mishler and Brandon advocate a version of the phylogenetic species concept. They require that all species form monophyletic units — lineages containing all and only the descendents of a common ancestor. Furthermore, Mishler and Brandon suggest that different evolutionary forces cause monophyletic lineages to be species taxa. Some phylogenetic species are maintained by interbreeding, others by ecological forces, and others by developmental constraints. The result is one classification of the tree of life, but different forces affect different branches on that tree. This form of pluralism is too conservative because it is wedded to a phylogenetic approach to species and rules out the possibility of interbreeding species that are not monophyletic. As we saw in the case of ancestral species, many non-monophyletic interbreeding species meet population

genetic standards for being cohesive evolutionary units. Mishler and Brandon's pluralism leaves out a significant kind of basal lineage on the tree of life.

3.3 Monist Responses

Needless to say, monists are not happy with pluralism. They offer many responses to species pluralism. One objection is that pluralism is an overly liberal approach to science [Sober, 1984; Ghiselin, 1987; Hull, 1987; 1999]. Hull and others ask how pluralists determine which species concepts, among the plurality of suggested concepts, should be accepted as legitimate. Should any concept proposed by a biologist be accepted? If pluralism offers no criteria for discerning among species concepts, then pluralism, according to Hull [1987], boils down to the position of 'anything goes.'

Ereshefsky [1992] and Dupré [1993] respond to this objection by suggesting criteria for accepting a proposed species concept. Those criteria are such epistemic virtues as empirical testability, internal consistency, and intertheorectic consistency. These are standard epistemic virtues that scientists and philosophers use for judging theories. In judging a species concept, one might ask, for example, if the concept's theoretical assumptions are testable. BSC, for instance, assumes that interbreeding is an important factor in maintaining the existence of stable groups of organisms. Whether interbreeding causes such stability is empirically testable, and biologists do test that hypothesis in the field and in the lab [Templeton, 1989]. Pluralists do not subscribe to a position of 'anything goes.' Species pluralists believe that legitimate species concepts must meet specific epistemic standards.

A recent response to species pluralism is that all well-accepted species concepts are captured by a more inclusive species concept. De Queiroz [1999] and Mayden [2002] observe that all prominent concepts assume that species taxa are lineages. Consequently, de Queiroz and Mayden offer a Lineage Concept of Species that, according to Mayden [2002, 191], "serves as the logical and fundamental over-arching conceptualization of what scientists hope to discover in nature behaving as species. As such, this concept... can be argued to serve as the primary concept of diversity." De Queiroz and Mayden believe that the species concepts currently accepted describe different types of species lineages — for example, interbreeding lineages, ecological lineages, and phylogenetic lineages, and the Lineage Species Concept provides the proper account of all species lineages. De Queiroz and Mayden believe that their species concept is a monist answer to pluralism because it provides the one correct account of species.

A problem with de Queiroz and Mayden's Lineage Species Concept is that it fails to highlight a unique group of entities that should be called 'species.' Recall that Mayden believes that the Lineage Species Concept captures that "over-arching conceptualization of what scientists hope to discover in nature behaving as species" (ibid.). Such a conceptualization should capture what is common and *unique* to all species taxa. Interbreeding, ecological, and phylogenetic species are all lineages,

so the Lineage Species Concept captures an important similarity of species taxa. However, all genera, families, and other Linnaean taxa are also lineages. The Lineage Concept is too inclusive because it captures all Linnaean taxa. In an attempt to provide an over-arching conceptualization of species taxa, de Queiroz and Mayden have cast their net too widely.

Consider a final monist response to pluralism inspired by advances in molecular genetic sequencing. Alex Rosenberg (in conversation) suggests that perhaps the correct species concept could be based on genetic similarity. As more molecular studies are performed, we may discover the distinctive genome of each species. We could then use that information to construct a single classification of the organic world. The problem with this suggestion is that classifying by molecular data would not unify biological taxonomy, instead it would add further classifications. As Ferguson [2002] observes, classifications based on overall genetic similarity and classifications based on the ability to interbreed do not coincide. The result is two different classifications: one classifying organisms by interbreeding, another classifying organisms by overall genetic similarity. The same sort of disunity is found when classifications based on overall genetic similarity and ecological adaptiveness are compared. Wu and Ting [2004] cite cases where classifications based on genes for ecological adpativeness fail to coincide with classifications based on overall genetic similarity. Bringing molecular data to the table does not reduce the number of classifications but increases their number and brings further disunity to biological taxonomy.

4 THE LINNAEAN HIERARCHY

The Linnaean hierarchy should be familiar to all readers of this volume. The hierarchy contains a series of ranks — species, genus, family, and so on — that serve as the primary categories for biological classification. The Linnaean ranks are also cited in biological theory. Prey-predator models in ecology refer to species, and hypotheses concerning the tempo and mode of macroevolution often refer to families, classes and more inclusive ranks. There is more to the Linnaean hierarchy than just a series of ranks. The Linnaean hierarchy also includes rules of nomenclature that tell us how to name taxa. For example, the Linnaean hierarchy tells us to give all and only species taxa binomial names, such as the name of our species, *Homo sapiens*.

Given the pivotal role of the Linnaean hierarchy in biology, one might be surprised to learn that a number of biologists and philosophers question the usefulness of that hierarchy. Critics of the Linnaean hierarchy offer a long list of problems [Hull, 1966; Hennig, 1969; de Queiroz and Gauthier, 1992; Ereshefsky, 2001]. Some critics believe that the problems facing the Linnaean hierarchy are so severe that the Linnaean hierarchy should be replaced with an alternative system of classification [Hennig, 1969; de Queiroz and Gauthier, 1992; Ereshefsky, 2001; Cantino *et al.*, 2003]. The Linnaean hierarchy has its defenders as well. Defenders are well aware of the problems facing the Linnaean hierarchy, but they argue that

alternative systems have their own problems [Mayr, 1969; Wiley, 1981; Forey, 2002; Brummit, 2002]. This section introduces the problems facing the Linnaean hierarchy as well as alternatives to that system.

4.1 Linnaean Ranks

Linnaeus's original hierarchy consisted of five ranks: subspecies (variety), species, genus, order and class. Taxonomists in the early 20^{th} Century found Linnaeus's 5 ranks insufficient for representing life's diversity, so they increased the number of ranks to 21. A persistent question concerning the Linnaean ranks is how to define them. We have seen that biologists and philosophers spend a considerable amount of time defining the species category. What of the other Linnaean ranks? How are they defined? Two schools of taxonomy dominated 20^{th} Century biology: Evolutionary Taxonomy and Cladism. Biologist in each school proposed different definitions of the higher Linnaean ranks (those ranks above the rank of species). However, none of those definitions have withstood criticism.

Before getting to those definitions we need a quick introduction to Evolutionary Taxonomy and Cladism. Evolutionary Taxonomy was founded by Ernst Mayr [1969] and Gaylord Simpson [1961]. Evolutionary taxonomists attempt to capture two types of phenomena in taxa. First, a taxon is a single genealogical entity. Second, the members of a taxon inhabit a common adaptive zone and share a common way of life. Birds, for example, form a taxon for evolutionary taxonomists because birds are a single genealogical entity, and birds, for the most part, share a common adaptive zone that causes them to have a relatively distinct way of life. The other major taxonomic school of the 20^{th} Century, Cladism, was founded by Willi Hennig [1966]. Cladists attempt to capture only one type of biological phenomena — genealogy. For cladists, a taxon contains all and only the descendents of a common ancestor. Such taxa are monophyletic. Cladists do not recognize birds as a taxon because doing so would make reptilia a non-monophyletic taxon. Crocodiles are more closely related to birds than to other reptiles, yet crocodiles are recognized as part of reptilia, see Figure 4.1. If birds are recognized as a taxon separate from reptilia, then reptilia would not be monophyletic — some of the reptilia's descendants, birds, would not be included in reptilia. Only genealogy matters for cladists, and removing birds from reptilia has no genealogical justification.

Because evolutionary taxonomists and cladists disagree on what counts as a taxon, they disagree on how to define the higher Linnaean ranks. For the evolutionary taxonomist, the higher Linnaean ranks are defined in terms of phenotypic diversity and ecological breadth: the greater the phenotypic diversity and ecological breadth of a taxon, the more inclusive the taxon. All families, for example, contain a certain degree of phenotypic diversity; and, all families occupy an adaptive zone of a certain width. The adaptive zone of a family will be smaller than the adaptive zone of a tribe, which is why, according to evolutionary taxonomists, families are less inclusive taxa than tribes.

The concepts of 'phenotypic diversity' and 'adaptive zone' were widely used

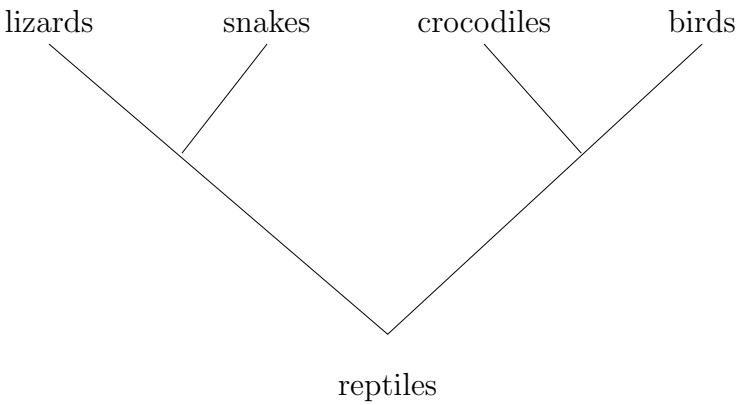


Figure 2. Reptilia is paraphyletic because it contains lizards, snakes, and crocodiles but not birds.

in the 20^{th} Century. But now many taxonomists object to their use in determining a taxon's Linnaean rank [Eldredge and Cracraft, 1980]. They argue that the concepts of 'phenotypic diversity' and 'adaptive zone' are imprecise and used ambiguously. What counts as a distinct adaptive zone varies from phylum to phylum. When it comes to phenotypic diversity, Hennig [1966, 156] playfully asks, "whether the morphological divergence between an earthworm and a lion is more or less than between a snail and a chimpanzee?" Most taxonomists now believe that the concepts of 'phenotypic diversity' and 'adaptive zone' are too malleable to serve as measures of a taxon's rank.

Hennig [1965) offered his own definitions of the higher Linnaean ranks. He suggested that taxa of the same rank originate during the same time period. Just as geological strata are organized according to time of origin, biological taxa should be assigned Linnaean ranks according to their time of origin. All taxa assigned the rank of class, for example, are taxa that originated during the Late Cretaceous. Orders would be defined as those taxa that originated during a more recent time. Hennig's suggested definitions of the Linnaean ranks are more precise than the definitions offered by evolutionary taxonomists. Unfortunately, Hennig's definitions are problematic from a cladistic perspective. (Recall that Hennig is the founder of the taxonomic school cladism.) Taxa that originate during the same period often have different phylogenetic or branching structures. Some taxa originating during the Late Cretaceous are very successful and contain a number of subtaxa. Other taxa originating during the same time are monotypic and contain only a single basal taxon: they are phylogenetic twigs. From a cladistic perspective, Hennig's definitions place different types of taxa under a single rank. Consequently, cladists, including later Hennig, abandoned the idea of correlating the ranks of taxa with times of origin.

Neither evolutionary taxonomists nor cladists have established a universal cri-

terion for defining the higher Linnaean ranks. Instead, biologists use a patchwork of criteria for defining the higher ranks. As a result, taxa of the same rank can vary dramatically. Families can vary in their age, their phylogenetic structure, their phenotypic diversity, and the breadth of their adaptive zone. Calling a taxon a 'family' merely means that within a particular classification that taxon is more inclusive than a genus and less inclusive than a class. There is no definition for 'family' that applies to all or even most classifications of families. 'Family,' like the other Linnaean ranks, refers to a heterogeneous set of taxa. Some authors have suggested that the Linnaean hierarchy of ranks is a fictitious grid we place on nature [de Queiroz and Gauthier, 1992; Ereshefsky, 1994].

The heterogeneity of the Linnaean ranks has practical implications. Consider the use of Linnaean ranks in biodiversity studies. Biologists tend to measure biodiversity in Linnaean terms; an area or higher taxon is surveyed for the number of species or families present. However, the Linnaean ranks mask important biological differences. Suppose we want to measure the biodiversity of two groups of organisms and we measure that diversity by numbers of families. One group consists of snail families, and the other contains mammalian families. Snail families and mammalian families are biologically different. Snail families have denser phylogenetic structures than mammalian families; and, mammalian families exhibit more ecological diversity than snail families. Snail and mammalian families are not comparable. Counting biodiversity by the number of families present in a group masks important biological differences. Biodiversity studies should instead use parameters that capture such biological phenomena as phylogenetic structure or ecological breadth. Then we would have proper measures of biodiversity.

4.2 Linnaean Names

As mentioned earlier, the Linnaean hierarchy is a general system of classification containing both hierarchical ranks and rules for naming taxa. The Linnaean ranks are not the only aspect of the Linnaean hierarchy that is problematic, so are the Linnaean rules of nomenclature. Critics believe that the Linnaean rules of nomenclature lead to a number of practical problems in taxonomy [de Queiroz and Gauthier, 1992; Ereshefsky, 2001]. What follows is a quick introduction to those problems.

Linnaeus's best-known naming rule is his binomial rule. Each species' name has two parts: a generic name and a specific name. In *Homo sapiens*, *Homo* is the name of our genus and *sapiens* is the specific name of our species. Binomials clearly indicate the classification of a species: *Homo* is the name of our species' genus. *Sapiens*, the specific name, distinguishes our species from other species in *Homo*. Linnaeus's motivation for assigning binomial names was his belief that a biologist should memorize the classification of all species in a kingdom. Linnaeus realized that there were too many species to do that; for example, he recognized approximately 10,000 plant species. He believed that the number of plant genera in the world was much smaller, approximately 300 genera. Furthermore, he did

not think that number would greatly increase given his assumption of how much of the world had been explored. With not too much difficulty, a biologist could memorize all the names and taxonomic positions of genera in a kingdom. Once those names and positions were memorized, a biologist would know the classification of each species in a kingdom by reading the generic name in a species' binomial name. Linnaeus's binomials, thus, served as guides for memorizing the taxonomic positions of species in a kingdom.

Binomials may have served their intended purpose in Linnaeus's time when it was assumed that there was approximately 300 plant genera and 300 animal genera and those numbers would not increase greatly. The problem is that Linnaeus did not envision the extent of biodiversity in the world, and so he grossly underestimated the number of genera in the world. Conservative estimates put the numbers of plant genera in the tens of thousands. Estimates for animal genera are in the tens of thousands as well. Given these numbers, there are too many generic names in a kingdom to memorize. As Mayr [1969, 334] writes, "a generic name no longer tells much to a zoologist except in a few popular groups of animals." Binomials have lost their original motivation: they do not serve as guides for memorizing the classification of all the species in a kingdom. Nevertheless, biologists are still required to assign a species both a generic and a specific name. The binomial rule remains in place even though it fails to achieve its original aim.

One might wonder if the continued use of the binomial rule is problematic. Though the original motivation is gone, there seems no harm in incorporating a generic name in a species name. However, the use of binomials has its costs. The binomial rule can place a biologist in the awkward position of having to assign a species to a genus before having the proper empirical information for making that assignment. The binomial rule, in other words, can lead to hasty classification. Recall that the binomial rule requires that a species be assigned to a genus before it can be named. But in some situations, a biologist lacks the proper information for assigning a species to a genus. In such situations, if a biologist wants to name a newly discovered species, she must assign that species to a genus on the basis of insufficient information. This is no small problem. Cain [1959, 242] writes that "the necessity of putting species into a genus before it can be named at all is responsible for the fact that a great deal of uncertainty is wholly cloaked and concealed in modern classification." Thus, the use of the binomial rule is not so innocuous. It causes biologists to assign species to genera without having sufficient data.

There are further problems with the Linnaean rules of nomenclature, and these involve not just Linnaeus's binomial rule but also other Linnaean rules. Many of these rules of nomenclature were introduced by taxonomists in the 20^{th} Century. For example, one rule is that the names of taxa have rank-specific endings that indicate the Linnaean rank of a taxon. For example, the suffix — *idae* in the names 'Hominidae' and 'Tipulidae' indicate that such taxa are families. The suffix — *ini* shows that Hominini is a tribe. The inclusion of rank-specific endings in the names of higher taxa causes further problems with the Linnaean hierarchy.

Consider the case of taxonomic revision. Taxonomic revision is the activity of revising a classification. Such revision is the norm not the exception in biological taxonomy. A taxon may be reassigned to another higher taxon, for example, a species may be reassigned to another genus. Or revision may involve giving a taxon a new rank, for example, a taxon thought to be a tribe is reassigned as a family. Taxonomic revision occurs for a couple of reasons. New empirical information may be gathered, such as new DNA evidence or the discovery of new fossils, that causes revision. Or revision may occur when taxonomic theory changes and old classifications are updated to cohere with the new theory. Taxonomic revision causes instability in classification, and this is to be expected. Classifications are hypotheses and are open to revision in light of new evidence or theoretical considerations. Unfortunately, the Linnaean rules of nomenclature are themselves a cause of instability because they require that the names of taxa be changed when revision occurs. Recall that in the Linnaean hierarchy, the names of taxa reflect a taxon's rank and taxonomic position. A change in a taxon's classification requires a change in a taxon's name. This may not sound like much of an inconvenience, but it is. A case of taxonomic revision can involve the renaming of hundreds of taxa (Ereshefsky 2001). The Linnaean rules themselves are a source of instability.

Another area where the Linnaean rules of nomenclature make classifying taxa harder than need be is that of taxonomic disagreement. In some cases biologists disagree over the classification of a taxon. The disagreement may be over the rank of a taxon, or the disagreement may be over the taxonomic placement of a taxon (does it belong to this or that more inclusive taxon). The Linnaean rules require that the rank and placement of a taxon be reflected in its name. If biologists disagree over a taxon's rank, for instance, the Linnaean rules require those biologists to assign different names to what they agree is the same taxon. Consider a disagreement between E.O. Wiley and G. G. Simpson concerning the placement of our genus, *Homo*. They agree that *Homo* belongs to a more inclusive taxon, call it 'X,' but they disagree on X's Linnaean rank. They agree on the members and properties of X, but because they belong to different taxonomic schools they assign X different ranks. Wiley thinks X is a tribe. Simpson thinks it is a family. Following the Linnaean rules, Wiley calls X *Hominini*, and Simpson calls X *Hominidae*. The Linnaean rules force Simpson and Wiley to give different names to what they agree is the same taxon. The Linnaean rules are, thus, a source of semantic confusion.

4.3 Alternatives to the Linnaean Hierarchy

In light of the problems facing the Linnaean hierarchy, several alternative systems of classification have been suggested in the last 40 years. Hull [1966] and Hennig [1969] offered alternative systems, and more recently a system of phylogenetic nomenclature called the "PhyloCode" has been suggested [de Queiroz and Gauthier, 1992; Cantino *et al.*, 2003. All of these alternatives to the Linnaean hierarchy recommend eliminating Linnaean ranks from biological classification. They

promote a 'rankless taxonomy.' They offer a couple of alternative methods for indicating the hierarchical relations among taxa. Here we will consider just one suggested method.

According to Hennig [1969], Linnaean ranks should be replaced with positional numbers. Consider a standard Linnaean classification.

Subclass Reptilomorpha
Infraclass Aves
Infraclass Mammalia
Division Monotremata
Division Theria
Cohort Metaheria
Cohort Eutheria.

A classification of the same taxa using Hennig's positional numbers is the following:

2.4.	Reptilomorpha
2.4.1.	Aves
2.4.2.	Mammalia
2.4.2.1	Monotremata
2.4.2.2.	Theria
2.4.2.2.1.	Metaheria
2.4.2.2.2.	Eutheria.

Positional numbers indicate a couple of things. First, positional numbers indicate hierarchical relations among taxa. Eurtheria is a part of Mammalia and this is shown by Eurtheria's positional number containing the positional number of the more inclusive Mammalia. Second, positional numbers indicate the degree of inclusiveness of a taxon: the fewer the digits in a taxon's positional number, the more inclusive the taxon.

Hennig's positional number system overcomes the problems associated with Linnaean ranks and names. Recall that a problem with the Linnaean ranks is the assumption that taxa of the same Linnaean rank are comparable across classifications (Section 4.1). Positional numbers carry no such assumption; positional numbers are merely notational devices to indicate the hierarchical relations of taxa within specific classifications. Positional numbers have no meaning outside of particular classifications. Consequently, no suspect ontological categories are associated with positional numbers, as is the case with the Linnaean ranks.

Another feature of Hennig's positional number system is that it avoids the naming problems facing the Linnaean rules of nomenclature. Recall that under the Linnaean rules, a taxon's name is not merely a name but also a device for indicating a taxon's classification. Taxon names in Hennig's system do not play that dual role: taxon names are merely names; positional numbers perform the task of indicating a taxon's placement in a classification. In Hennig's system, the activity of naming taxa is divorced from the activity of classifying taxa. By keeping

the activities of naming and classifying separate, Hennig's system overcomes the naming problems highlighted earlier.

Consider the affect taxonomic revision has on Linnaean names. In the Linnaean hierarchy, a name indicates a taxon's classification, so a change in classification requires a change in name. In the positional system, taxonomic revision requires a change in a taxon's positional number, but the name of the taxon remains the same. Unlike the Linnaean rules, names remain stable during taxonomic revision. The positional system also avoids the problem posed by taxonomic disagreement. In the Linnaean hierarchy, when two biologists disagree on the rank of a taxon, they must assign that taxon different names, each indicating the different ranks assigned to the taxon. With positional numbers, a taxon has a single name, and biologists display their disagreement over that taxon's placement by assigning that taxon different positional numbers. Hence, the use of positional number avoids the semantic confusion caused by the Linnaean rules. Finally, recall the problem of hasty classification caused by the Linnaean binomial rule. In the Linnaean hierarchy, a biologist must first determine the genus of a species before she can name it, even though she may lack adequate empirical evidence for assigning a species to a genus. In Hennig's system, a species can be named before knowing that species' classification; a name is assigned to the species, and its positional number is determined later.

The positional number system nicely overcomes the ranking and naming problems outlined above. However, some proponents of non-Linnaean systems think that further changes in nomenclature are necessary. For example, a current controversy among proponents of non-Linnaean systems is what to do with the names of species. Some post-Linnaeans suggest converting species binomial to uninomials (Cantino et al. 1999). There are two motivations for doing so. One is the belief that because there is no species category in nature, those taxa we call 'species' should not be given a special notational device; all taxa should have uninomial names. Another motivation for eliminating binomials has to do with revision. Taxa with binomial names may be assigned to more inclusive positions in classifications. If that occurs, then taxa with binomial names occur at various hierarchical levels in a classification. Classifications would then have the confusing feature of binomials referring to non-basal taxa.

How, then, should binomials be converted to uninomials? According to one suggestion, a binomial should be converted to a uninomial by placing the specific and generic name together. The binomial *Canis Familaris* would become the uninomial 'Canisfamilaris.' According to another proposal, the generic name of a species should be dropped and a registration number should be added to the specific name. *Canis Familaris* would become 'Familaris5732,' for example. The registration number is added to avoid homonyms.

4.4 A Middle Ground

Defenders of the Linnaean hierarchy worry that the proposed alternative systems are too radical [Forey, 2002; Brummitt, 2002]. They worry that the replacing the Linnaean hierarchy would be overly disruptive to biological practice. For example, suppose that the Linnaean ranks were replaced by positional numbers. The Linnaean ranks are well entrenched both in and outside of biology. The ranks of species and genera occur in text books, field guides, environmental legislation, and elsewhere. Arguably, replacing the Linnaean ranks both in and outside of biology would be too disruptive. A similar case is made against changing the names of taxa, for example, replacing binomial names with uninomials. Switching binomials to uninomials would require rewriting countless classifications, textbooks, and field guides. Critics of non-Linnaean systems think that changing the names of taxa would be too disruptive and impractical.

Still, the Linnaean hierarchy has its problems, as illustrated in Sections 4.1 and 4.2. Keeping the Linnaean hierarchy in place avoids disrupting biological practice, but it also keeps in place the problems caused by the Linnaean hierarchy. Given the pros and cons of replacing the Linnaean hierarchy, what should be done? There is a middle ground between replacing the Linnaean hierarchy and keeping the Linnaean hierarchy as it stands. We could keep the Linnaean ranks and names as they are, but rid them of their Linnaean meaning.

When it comes to the names of taxa, we could keep the current taxon names yet deny that such names having any classificatory meaning. A binomial name, for example, would just be a name, it would no longer indicate the rank of a taxon. Thus during taxonomic revision a taxon with a binomial name would keep its name even if that taxon were reclassified as a more inclusive taxon. Similarly, two biologists that disagreed on the rank of a taxon could continue to use its binomial name because that name has no classificatory meaning — it is merely a name. Keeping taxon names constant but eliminating any classificatory meaning associated with such names avoids the problems facing the Linnaean rules while at the same time keeping names stable.

A similar approach can be applied to the Linnaean ranks. We can keep the Linnaean ranks and restrict the meaning of those ranks to being indicators of the hierarchical relations among taxa within a *particular* classification. At the same time, we would rid the Linnaean ranks of their metaphysical connotations: the Linnaean ranks would not highlight any categories in nature; and, the assumption that all taxa of the same rank, for example, all genera, share an ontological similarity would be dropped. This approach avoids the problems associated with the Linnaean ranks. For example, false comparisons among taxa of the same Linnaean rank in biodiversity studies would be eliminated. At the same time, the Linnaean ranks would still remain in place, so the disruption that would be caused by junking the Linnaean ranks is avoided. This approach to Linnaean nomenclature and ranks charts a middle ground between critics and defenders of the Linnaean hierarchy. It acknowledges and avoids the problems facing the Linnaean hierarchy

while at the same time keeping biological taxonomy as stable as possible.[2]

Stepping back, in this chapter we have seen a variety of conceptual issues at the forefront of biological taxonomy and systematics. Species may not be qualitative kinds but historical entities defined by genealogy. The idea of single correct classification of the organic world may need to be replaced with a more pluralistic approach to classification. The Linnaean ranks that pervade biology may have no basis in nature but are merely instruments for organizing life's diversity. Other radical changes in how we classify and bring order to the organic world may be just around the corner. For example, life on this planet may not be a single genealogical tree, but a tangled bush [Doolittle, 1999]. If life is a tangled bush, then hierarchical classifications may be the wrong way to represent life's diversity. Biological systematics is rife with conceptual issues in need of philosophical analysis.

BIBLIOGRAPHY

[Boyd, 1999a] R. Boyd. Homeostasis, species, and higher taxa. In R. Wilson (ed.), *Species: New Interdisciplinary Essays*, 141–185. Cambridge: MIT Press, 1999.

[Boyd, 1999b] R. Boyd. Kinds, complexity and multiple realization: comments on Millikan's "Historical Kinds and the Special Sciences". *Philosophical Studies* 95, 67–98, 1999.

[Brummitt, 2002] R. Brummitt. How to chop up a tree. *Taxon* 51, 31-41, 2002.

[Cain, 1959] A. Cain. The Post-Linnaean Development of Taxonomy. *Proceedings of the Linnaean Society of London* 170:234-244, 1959

[Cantino, 1999] P. D. Cantino, *et al.*. Species names in phylogenetic nomenclature. *Systematic Biology* 48: 790-807, 1999.

[Cantino, 2003] P. D. Cantino, *et al.*. Phylocode: a phylogenetic code of biological nomenclature, 2003. `http:/www.ohiou.edu/phylocode/`

[Claridge *et al.*, 1997] M. F. Claridge, A. Dawah, and M. R. Wilson, eds. *Species: The Units of Biodiversity*. London: Chapman and Hall, 1997.

[de Queiroz and Gauthier, 1992] K. de Queiroz and J. Gauthier. Phylogenetic taxonomy. *Annual Review of Ecology and Systematics* 23, 480–499, 1992.

[de Queiroz, 1999] K. de Queiroz. The general lineage concept of species and the defining properties of the species category. In R. Wilson (ed.), *Species: New Interdisciplinary Essays*, 49–90. Cambridge: MIT Press, 1999.

[Doolittle, 1999] W. F. Doolittle. Phylogenetic classification and the universal tree. *Science* 284 (June), 2124–2128, 1999.

[Dupré, 1981] J. Dupré. Natural kinds and biological taxa. *Philosophical Review* 90, 66–90, 1981.

[Dupré, 1993] J. Dupré. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge: Harvard University Press, 1993.

[Eldridge and Cracraft, 1980] N. Eldridge and J. Cracraft. *Phylogenetic Patterns and the Evolutionary Process*. New York: Columbia University Press, 1980.

[Ereshefsky, 1991] M. Ereshefsky. Species, Higher Taxa, and the Units of Evolution. *Philosophy of Science*. 58: 84-101, 1991.

[Ereshefsky, 1992] M. Ereshefsky. Eliminative Pluralism. *Philosophy of Science* 59, 671- 90, 1992.

[2]Wiley [1981] also suggests a revised Linnaean hiearchy, what he calls the "annotated Linnaean hierarchy." Wiley's annotated Linnaean hierarchy is different than the revised Linnaean hierarchy offered here. To cite one difference, Wiley's revision does not divorce the activity of naming taxa from the activity of classifying taxa. Thus, Wiley's annotated Linnaean hierarchy does not address or solve any of the naming problems mentioned in Section 4.2.

[Ereshefsky, 1994] M. Ereshefsky. Some problems with the Linnaean hierarchy. *Philosophy of Science* 61, 186–205, 1994.
[Ereshefsky, 2001] M. Ereshefsky. *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge: Cambridge University Press, 2001.
[Ereshefsky and Matthen, 2005] M. Ereshefsky and M. Matthen. Taxonomy, polymorphism and history: an introduction to population structure theory, 2005.
[Ferguson, 2002] Ferguson, J. 2002: On the use of genetic divergence for identifying species. *Biological Journal of the Linnean Society* 75, 509–519, 2002.
[Forey, 2002] P. Forey. Phylocode — pain, no gain. *Taxon* 51, 43–54, 2002.
[Ghiselin, 1974] M. Ghiselin. A radical solution to the species problem. *Systematic Zoology* 23, 536–544, 1974.
[Griffiths, 1999] P. Griffiths. Squaring the circle: natural kinds with historical essences. In R. Wilson (ed.), *Species: New Interdisciplinary Essays*, 209–228. Cambridge: MIT Press, 1999.
[Hennig, 1965] W. Hennig. Phylogenetic systematics. *Annual Review of Entomology* 10, 97–116, 1965.
[Hennig, 1966] W. Hennig. *Phylogenetic Systematics*. Urbana: University of Illinois Press, 1966.
[Henning, 1969] W. Hennig. *Insect Phylogeny*. John Wiley Press, New York, 1969.
[Hull, 1965] D. Hull. The effect of essentialism on taxonomy: two thousand years of stasis. *British Journal for the Philosophy of Science* 15, 314–326, 1965.
[Hull, 1966] D. Hull. Phylogenetic Numericlature. *Systematic Zoology* 15:14-17, 1966.
[Hull, 1978] D. Hull. A matter of individuality. *Philosophy of Science* 45, 333–360, 1978.
[Hull, 1987] D. Hull. Genealogical Actors in Ecological Roles. *Biology and Philosophy* 2: 168–183, 1987.
[Hull, 1999] D. Hull. On the Plurality of Species: Questioning the Party Line. In R. Wilson (ed.), *Species: New Interdisciplinary Essays*, 23 — 48. Cambridge: MIT Press, 1999.
[Kitcher, 1984] P. S. Kitcher. Species. *Philosophy of Science* 51, 308–333, 1984.
[Kripke, 1982] S. Kripke. Naming and Necessity. In D. Davidson and G. Harmon (eds.), *Semantics of Natural Language*, 253-355. Reidel: Dordrecht, 1982.
[Mayden, 2002] R. Mayden. On biological species, species concepts and individuation in the natural world. *Fish and Fisheries* 3, 171–196, 2002.
[Mayr, 1959] E. Mayr. Typological versus Population Thinking. *In Evolution and Anthropology: A Centennial Appraisal*. The Anthropological Society of Washington, Washington D.C., 1959
[Mayr, 1969] E. Mayr. *Principles of Systematic Zoology*. Harvard University Press, Cambridge, Massachusetts, 1969.
[Mayr, 1970] E. Mayr. *Populations, Species, and Evolution*. Cambridge, Massachusetts: Harvard University Press, 1970.
[Mayr, 1982] E. Mayr. *The Growth of Biological Thought*. Cambridge, Massachusetts: Harvard University Press, 1982.
[Millikan, 1999] R. Millikan. Historical Kinds and the "Special Sciences". *Philosophical Studies* 95: 45-65, 1999.
[Mishler and Bramdon, 1987] B. Mishler and R. Brandon. Individuality, Pluralism, and the Phylogenetic Species Concept. *Biology and Philosophy*, 2: 397-414, 1987.
[Putnam, 1975] H. Putnam. *Mind Language, and Reality. Philosophical Papers,* Volume 2. Cambridge University Press, Cambridge, 1975.
[Rosenberg, 1994] A. Rosenberg. Instrumental Biology or the Disunity of Science, Chicago: Chicago University Pressm 1994.
[Ruse, 1987] M. Ruse. Biological Species: Natural Kinds, Individuals, or What? *British Journal for the Philosophy of Science* 38: 225-242, 1987.
[Simpson, 1961] G. Simpson. *The Principles of Animal Taxonomy*. Columbia University Press, New York, 1961.
[Sober, 1980] E. Sober. Evolution, population thinking and essentialism. *Philosophy of Science* 47, 350–383, 1980.
[Sober, 1984] E. Sober. Sets, Species, and Natural Kinds. *Philosophy of Science* 51, 334-41, 1984.
[van Valen, 1976] L. 0Van Valen. Ecological Species, Multispecies, and Oaks. *Taxon*: 25:233-239, 1976.
[Wiley, 1981] E. Wiley. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. Wiley and Sons, New York, 1981.

[Wilson, 1968] E. O. Wilson. The Ergonomics of Caste in the Social Insects. *American Naturalist* 102: 41-66, 1968.

[Wilson, 1999] R. Wilson. Realism, essence, and kind: resuscitating species essentialism? In R. Wilson (Ed.), *Species: New Interdisciplinary Essays*, 187 — 208. Cambridge: MIT Press, 1999.

[Winsor, 2003] M. P. Winsor. The Non-Essentialist Methods in Pre-Darwinian Taxonomy. *Biology and Philosophy* 18: 387-400, 2003.

[Wu and Ting, 2004] C. Wu and C. Ting. Genes and speciation. *Nature Genetics* 5, 114 –122, 2004.

This page intentionally left blank

HOMOLOGY AND HOMOPLASY

Brian K. Hall

> "In [the various] kinds of animals and plants [we see] simply the parts of one great genealogical tree, which have become detached and separated from one another in a thousand different degrees, through the operation of the great destroyer Time ..." [Lankester, 1870a, 34].

1 INTRODUCTION

Homology, from the Greek *Homologia*, agreement, has been foundational for any comparisons of biological objects for millennia. Homology normally is contrasted with analogy (similarity of function) and/or with homoplasy (*similarity* arising through independent descent).[1]

Clasically, homology and analogy both referred to similar parts (features) of organisms.[2] Because homology applies across the entire biological hierarchy from genes to individuals — perhaps even to populations and communities ([Hall, 1994a,b; 2004; 2007a]; [Hall *et al.*, 2004]) — and even though the *definition* of homology is invariant across the hierarchy, the *application* of homology must specify the level of the biological hierarchy to which the term and concept is applied, and so we must speak of 'homologous as an appendage', 'homologous as a gene network', and so forth. Homology at the level of the phenotype (phenotypic or structural homology) is the continuous occurrence of the same feature (be it gene, gene network, cell type, tissue, organ, structure or behaviour) in two organisms whose common ancestor possessed the feature. The same definition applies to synapomorphy, which is a derived character, shared by two or more different organisms.[3] Operationally, knowledge of a homologue precedes use of that knowledge to identity a feature as a synapomorphy.

[1] See Hall [1994a] and Bock and Cardew [1999] for homology, Sanderson and Hufford [1996] for homoplasy, and Cohen [1994] and Panchen [1999] for analogy. My entry draws on material from Hall [1994a; 2003a] which may be consulted for fuller development and illustration of the arguments.

[2] See Moment [1945], Smith [1967], Bock and von Wahlert [1965], Gans [1985], Rieppel [1988], Bock [1991] and Hall [2003a; 2007a] for discussions and analyses of similarity and of features in extant and extinct organisms.

[3] A character is any trait or feature of the phenotype, for which see Patterson [1982], M. H. Wake [1996], Matthen [2000], P. J. Wagner [2000], McShea [2000], and the chapters in Wagner [2001] and in Hall and Olson [2003].

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

2 SIMILARITY

Homologous features need not be identical but must share sufficient 'similarity' to be recognizable as homologous. Analogous and homoplastic features also are identified on the basis of similarity of structure or function.[4]

Similarity does not necessitate structural equivalence; some features known to be homologous could not be identified as such on the basis of structural similarity but can be so identified because we know the phylogenetic history of the feature and of the taxa that contain the feature, a history that links the feature to homologous features in earlier taxa. The term and concept, *latent homology*, has been used for such features (Box 1).

Homology is similarity that reflects common descent and ancestry. *Homoplasy* is similarity (some might say superficial similarity) arrived at via independent evolution. It is often (usually?) assumed that homologous features share a common developmental basis. However, different environments or selective pressures can trigger the appearance of what appear to be similar features in organisms that do *not* share a most recent common ancestor — homoplasy as classically defined — and where the features therefore develop using different not similar developmental processes. Parallelism, traditionally one of the classes of homoplasy, *can* produce similar features using *similar* developmental processes. Furthermore, and there are many examples, similar (homologous) aspects of the phenotype can arise using different (nonhomologous) developmental pathways. Homology may or may not imply conserved development. Homoplasy implies divergent development. Consequently, divergence in developmental pathways is not an adequate criterion to establish features as homoplastic.[5]

Importantly, some features of all organisms share some degree of relationship and similarity to one another. Shared relationship ranges from: similarity or even identity of structure in related taxa (exemplified by the limbs of tetrapods), through similarity of a feature in organisms with a shared ancestor (exemplified by fish fins and tetrapod limbs), to similarity of a feature in organisms whose last common ancestor lies deep in metazoan history (exemplified by the tissue cartilage in invertebrates and vertebrates) and finally to similarity and a shared ancestor

[4]Determining similarity is not trivial. Gans [1985] provides a valuable evaluation of the interrelationships and potential independence of similarity of structure, development and function. Also see Haas and Simpson [1946], Riedl [1978], Bowler [1984; 1996], Stevens [1984], Sanderson and Hufford [1996], D. B. Wake [1991; 1994; 1996; 1999], Hall [1994a; 1999a; 2003a,b], Ospovat [1995], Bock and Cardew [1999], Butler and Saidel [2000] and Rieppel and Kearney [2002] for what is meant by similarity.

[5]See Spemann [1915], de Beer [1971], Riedl [1978], Hall [1994a,b; 1995], Bock and Cardew [1999] and Svensson [2004] for discussions and examples of divergent (non-homologous) developmental bases of homologues, Reilly and Lauder [1988] and Hall [2002a; 2003a; 2007a] for parallelism, and Hall [2003b; 2004] and Hall *et al.* [2004] for the role of the environment. Hall [1992] proposed the terms equivalent and non-equivalent developmental processes to reflect conserved or divergent development. I still prefer these terms to homologous and non-homologous [Striedter and Northcutt, 1991], if only to avoid confusion over levels of homology. Butler and Saidel [2000] proposed *syngeny* and *allogeny* for similar or different genesis.

deep within animal phylogeny (exemplified by the paired appendages of insects and vertebrates.[6]

3 HIERARCHY AND EMBRYONIC DEVELOPMENT

When considered from the perspective of embryonic development, phenotypic feature may share developmental pathways, share portions of pathways, or develop using divergent developmental pathways. Furthermore, developmental pathways may be shared (and so by implication, conserved) even when a feature is not fully formed but exists as a rudiment or vestige, or when a structure appears only in some individuals as an atavism (a feature not expressed in the previous generation). These apparent discontinuities or disjunctions occur — and a hierarchical approach to homology is required — because the evolution of processes and patterns (final form) need not be congruent.

Homology, therefore, must be approached *hierarchically.* Homology at one level — a feature such as a limb — need not correspond to homology at other levels — the developmental processes that produce the limb, or the genetic cascades underlying those processes. Hall [1995a] discussed mechanisms of gastrulation, the origin of germ cells, and the mode of induction of the lens of the eye and of Meckel's cartilage as examples of homologous features arising from non-homologous developmental process. Stone and Hall [2004] discussed the transition of fish fins to tetrapod limbs, anterior walking appendages to feeding appendages in crustaceans, the origin of insect wings and the evolution(s) of the mesoderm as examples that highlight the need for a hierarchical approach to any consideration of homology.[7]

4 RECONCEPTUALIZING HOMOLOGY AND HOMOPLASY

Homoplastic features (independent evolution) can share an affiliation (shared developmental processes) with homologous features. Can we find a common thread uniting developmental and phylogenetic approaches that might reveal a fundamental relationship between homology and homoplasy? Would that common thread constitute a *deep homology* of shared genetic, biochemical, cellular and developmental mechanisms? The sharing of regulatory genes, gene cascades, or gene networks is an important aspect of deep homology: homologues usually retain these shared bases, but need not; the processes underlying homoplastic features can diverge.[8]

[6]See Clack [2002] and Hall [2005a] for fins fish and tetrapod limbs, Cole and Hall [2004a,b] for cartilages in the Metazoa, and Shubin *et al.* [1997] and Carroll *et al.* [2005] for insect and vertebrate appendages.

[7]See de Beer [1938; 1971], Bolker [1992], Hall [1995a; 1998; 1999a] Minsuck and Keller [1996] and Stone and Hall [2004] for further examples, and see Freeman and Martindale [2002] and Martindale *et al.* [2004] for the origination of mesoderm.

[8]See McShea [1996], Shubin *et al.* [1997], Gerhart [2000], Hall [2002a,b; 2003a; 2007a] and Valentine [2004] for deep homology, and see Weiss [1994/1995], Abouheif [1997; 1999], Duboule

Homology and homoplasy do not represent a dichotomy of homology *vs.* homoplasy, homology *vs.* analogy, or homoplasy as non-homology. The position taken here — and developed more fully in Hall [2003a; 2007a] — is based upon a combined analysis of homology and homoplasy, embryos and adults, and nearness or distance of phylogenetic relationships, to demonstrate a continuum from homology through reversals, rudiments, vestiges, atavisms and parallelism — all of which constitute an expanded category of homology — to convergence as the single class of homoplasy. This reconceptualization takes an expanded homology as the foundation of comparative studies in biology at all levels from genes to phenotypes and as reflecting the descent with modification that has characterized and continues to characterize the evolution of the animal kingdom.[9] It turns out that the position is not new; Hall [2003a], and independently Gould [2002] traced it back to E. Ray Lankester's subdivision of homology into homology and homogeny in 1870.

5 HOMOLOGY AND ANALOGY DELINEATED

The roots of homologous thinking and the use of homologous features to order and classify organisms go back to Aristotle [Moore, 1986]. 'Modern' thinking on the topic is traced to Richard Owen, the first Superintendent (Director) of the British Museum (Natural History), who provided what remains our working definitions of a homologue (homology) and analogue (analogy): "Homologue... The same organ in different animals under every variety of form and function... Analogue... A part or organ in one animal which has the same function as another part or organ in a different animal" [Owen, 1843, 379, 374].[10]

Despite Darwin having changed fundamentally how we view homology, Owen's definition works surprisingly well today. Although after Darwin it could be stated that "a feature is homologous in two or more taxa if it can be traced back to the same feature in the presumptive common ancestor" [Mayr, 1982, 45, 232] the *criteria* for determining homologous structures remain today what they were in pre-Darwinian times, — similarity, position, connections. As a consequence, we can take Owen's definitions as current working/operational definitions and use them in the everyday comparison of features.

Despite this constancy, homology has generated more words on definitional matters than any other topic in biology with the possible exception of natural

and Wilkins [1998], Hall [1999a; 2003b], McGhee [2000], Newman and Müller [2000], Wilkins [2002] and Weiss and Buchanan [2004] for shared genes and gene networks.

[9]I do not consider plants but the same reconceptualization applies, as exemplified by Sattler [1984; 1992; 1994], Sattler and Jeune [1992] and Endress *et al.* [2000].

[10]Analyses of Owen's role may be found in Huxley [1894], Clark [1900], Spemann [1915], Boyden [1943; 1947], Ghiselin [1976], Richards [1987a], Rupke [1993; 1994], in the chapters in Hall ([1994a], especially that by Alec Panchen), Padian [1997], Panchen [1999], Matthen [2000], Camardi [2001] and Amundson [2007]. See Rieppel [1988], Hall [1994b] and Panchen [1999] for discussions of homology and analogy, and see Cohen [1994] for the same use of homology and analogy in the social sciences.

selection, where the issue is not definitional but more fundamental and related to the existence, mode of operation and role of natural selection. Mivart [1870] proposed twenty-five terms for subclasses of homology; Panchen [1992] listed nine under the transformational and taxic homology, while from his studies on plants, Sattler [1984] argued that we should replace the term homology with similarity, structural correspondence or structural relationship. Nineteen types are listed in the index to Hall [1994a], yet others in Haas and Simpson [1946] and Fitch [2000]. The man who provided the most rational subdivision of homology, E. Ray Lankester, proposed only a single new definition, but unlike so many others it embodied a new approach and a conceptual advance.

Karl Gegenbaur also converted to a post-Darwinian homology ([Szarski, 1949]; [Di Gregorio, 1995]). Gegenbaur [1859] considered homologies to be shared structural types. Eleven years later, having absorbed Darwin's *The Origin of Species*, Gegenbaur ([1870], translated 1878) had added common ancestry and common embryonic rudiments to his definition: "*das Verhltniss zwischen zwei Organen, die gleichen Abstammung besitzen somit aus der gleichen Anlage hervorgegangen sind"* [the relationship between two organs that share common origin (ancestry) and therefore were derived from the same anlagen (primordium)]. Lankester [1870a], like Gegenbaur, emphasized common ancestry. Indeed, Lankester favoured abandoning the term homology altogether, proposing in its place '*homogeny*' for similarity resulting from shared ancestry (see below).

5.1 Homology, Archetypes and Embryonic Development

Richard Owen's approach to homology was informed (but not straitjacketed) by embryonic development, especially the researches of Rathke [1839] and Huxley [1864] on the development of the skull of 'the' lamprey, considered to be the primitive vertebrate skull. Owen used these studies to develop his theory of the archetypal vertebrate skull and to extend the concept of the archetype as an abstracted or ideal form to the vertebral column and limbs and then to the entire vertebrated animal [Owen, 1846; 1848; 1849].[11] Owen's archetype was grounded solidly in embryology, specifically in the 'laws' of Karl Ernst von Baer (Box 2) as introduced into England by a Scottish physician, Martin Barry [1836-1837a,b] and an English physiologist, William Carpenter [1839].

Initially, Darwin was with Owen on the archetype, seeing Owen's archetype as *the* ancestral vertebrate. Darwin moved to an embryological — and therefore more mechanistic — view of the archetype. As he recounted to Huxley, "The discovery of the type or 'idea' (in *your* sense, for I detest the word as used by Owen, Agassiz

[11] The term archetype was introduced simultaneously and independently by Owen [1846] and by the London anatomist, Joseph Maclise [1846], both of whom saw identification of the archetype as the primary aim of comparative anatomy. For recent analyses see Desmond [1982], Hall [1992; 1994a], Rupke [1993; 1994], Bowler [1996] and Padian [1997]. For a modern version of the necessity of an archetypal concept in morphology, and more specifically in homology, see Young [1993]. For the development of a science of morphology, see M. H. Wake [2001], Wake and Summers [2003] and Hall [2005c].

& Co) of each great class, I cannot doubt is one of the very highest ends of Natural History... I sh^d [should] have thought that the archetype in imagination was always in some degree embryonic, & therefore capable of *generally undergoing* further development."[12] In seeing the archetype as embryonic and as capable of change (adaptable), Darwin also was influenced by the new laws of embryology proposed by von Baer (Box 2). Consequently, although not extensively used in *The Origin*, embryological evidence provided Darwin with powerful evidence for descent with modification.

Although Darwin embraced von Baer and the findings that were emerging on the stability of germ layers from which structures arose — and which provided a powerful basis for homology assessment [Hall, 1998] — Darwin was *not* prepared to cede primacy to community in embryonic structure as the ultimate arbiter of homologous relationships. "Thus, community in embryonic structure reveals community of descent; but dissimilarity in embryonic development does not prove discommunity of descent, for in one of two groups the developmental stages may have been suppressed, or may have been so greatly modified through adaptation to new habits of life, as to be no longer recognizable" [Darwin, 1910, 371–372]. Darwin recognized that develop evolves, something that many later workers forgot, ignored, or simply did not know.

Owen had taken the same position in 1846 when he stated that homology "is mainly, if not wholly, determined by the relative position and connection of the parts, and may exist independently of... similarity of development," and continued (admittedly, 162 pages later) that "there exists doubtless a close general resemblance in the mode of development of homologous parts; but this is subject to modification, like the forms, proportions, functions and very substance of such parts, without their essential homological relationships being thereby obliterated" [Owen, 1846, 6, 174]. Thus, it is important to recognize that embryonic development was not a criterion for homology from the very outset, taking Owen's definitions of 1843 as the outset.

Not all stood with Owen and Darwin, however. Haeckel [1866] concluded that homologous features reflect common descent. Many ever since expect homologues to arise using similar developmental processes, for example, "homology has come to signify an agreement in evolutionary derivation and in embryonic development" (cited in [Hubbs, 1944, 305]).[13]

[12]Letter from Charles Darwin to Thomas Henry Huxley, written on 23 April 1853, cited from Darwin and Seward [1903, Vol. 1, 73]. See Ospovat [1995], Hall [1992; 1994a; 1999a], Bowler [1996] and Richards [1987b] for Darwin's view of archetypes.

[13]See Moment [1945] and Hall [1994a; 1995a; 1999a] for homology and developmental processes, and Nyhart [1995], Bowler [1996], Hall [1999a; 2000a; 2007b] and Guralnick [2002] for homology and evolutionary embryology in the late 19^{th} C.

6 HOMOPLASY

Another way of comparing and classifying features among organisms is *homoplasy*, a term introduced by Lankester [1870] to incorporate an evolutionary dimension into Owen's definition. A pioneer of evolutionary morphology, especially analyses of relationships between germ layers, homology and classification, Lankester is perhaps better known for his emphasis on degeneration as an evolutionary force, as editor, reviser and translator of Haeckel's *History of Creation* [1876], and as the author of a series of books of essays on science (*Fireside Science; The Kingdom of Man; Extinct Animals*) aimed at the general reader. Lankester [1870a] dealt specifically with whether Owen's term analogy could be used for homoplasy, argued that analogy has a wider significance than given to it by Owen and saw that the term applied to features irrespective of whether their bearers shared a common ancestry: "Any two organs having the same functions are analogous, whether closely resembling each other in their structure and relation to other parts or not; and it is well to retain the word in that wide sense" (p. 41).[14]

7 HOMOGENY AND HOMOPLASY

Both Lankester and Gegenbaur placed homology into an evolutionary framework because they were staunch Darwinians: "in [the various] kinds of animals and plants [we see] simply the parts of one great genealogical tree, which have become detached and separated from one another in a thousand different degrees, through the operation of the great destroyer Time..." [Lankester, 1870a, 34]. Lankester was concerned that although the majority of evolutionists would agree that organs A and B were homologous in animals α and β because a common ancestor possessed the same organ, the term homology made no reference to evolutionary lineage, indeed it was typological, referring homologues to some ideal type. Consequently, Lankester thought it "necessary to have two terms in place of the one "homologue" [*ibid.* p. 36]. His criterion for establishing the *two classes of homology* was the evolutionary history of the organisms concerned. Lankester coined *homogeny* for similar features shared by two organisms as a consequence of common descent:

> Structures which are genetically related, in so far as they have a single representative in a common ancestor, may be called *homogenous*. We may trace an *homogeny* between them, and speak of one as the *homogen* of the other... details not traceable to, and inherited from the ancestor cannot be homogenous. [Lankester, 1870a, 36, his emphases]

Lankester introduced *homoplasy* for what he considered to be a single class of evolutionary phenomena. Homology and homoplasy *both* referred to the consequences of the actions of identical or nearly similar forces or environments, in the

[14] See Rieppel [1988], Hall [1994b; 2003a], Panchen [1999] and Gould [2002] for discussions of Lankester's position.

one case, acting on *two or more parts of an organism* (what had been called from Owen on, serial homology), in the other acting on *parts in two organisms*, the parts being exactly or nearly alike: "Homoplasy includes all cases of close resemblance of form which are not traceable to homogeny, all *details* of agreement not homogenous in structures which are broadly homogenous, as well as in structures having no genetic affinity [i.e., no connection through descent]" [Lankester, 1870a, 41].

Lankester made it very clear that homogeny and homoplasy were *two classes of homology*: "What is put forward here is this, — that under the term "homology", belonging to another philosophy, evolutionists have described and do describe two kinds of agreement — the one, now proposed to be called "homogeny", depending simply on the inheritance of a common part, the other, proposed to be called "homoplasy", depending on a common action of evoking causes or moulding environment on such homogenous parts, or on parts which for other reasons offer a likeness of material to begin with" [Lankester, 1870a, 42]. Equally clear was the reason for separating the two: "Darwinian morphology has further rendered necessary the introduction of the terms "homoplasy" and "homoplastic" to express that close agreement in *form* which may be attained in the course of evolutionary changes by organs or parts in two animals which have been subjected to similar moulding conditions of the environment, but have not close genetic community of origin [ancestry], to account for their similarity in form and structure" [Lankester, 1911, vol. 28, 1029].

Lankester's term homogeny did not take hold.[15] Homoplasy — similarity resulting from evolutionary convergence, parallelism or reversal — has endured, as exemplified in the book *Homoplasy: The Recurrence of Similarity in Evolution*, edited by Sanderson and Hufford [1996], produced under the expert editorial direction of Charles (Chuck) Crumly of Academic Press to parallel (no pun intended) *Homology: The Hierarchical Basis of Comparative Biology*, edited by Hall [1994a].

Adverse reaction to homogeny was rapid, especially from St. George Mivart who responded in the very next issue of *The Annals and Magazine of Natural History*, the journal in which Lankester had published his paper. Lankester [1870b] responded in the same issue; see Hall [2003a] for details. Mivart thought that abandoning the term homology would be "prejudicial to science", that "it is quite possible to have, on the one hand, developmental homogeny between parts which are not ancestrally homogenous, and, on the other, to have ancestral homogeny between parts which are not developmental homogenous" [Mivart, 1870, 116], and proposed 25 different types of homology to support his arguments. A year

[15]Lankester introduced the term *homotrophic* for what Darwin [1910] had discussed as correlations, compensation and economy of growth, which, for Lankester [1870a, 39] reflected the "delicate balancing of the forces of the organism, which would cause the disturbance of equilibrium in one part to affect simultaneously another part equally and similarly. Organs which stand in this nutritional relation to one another may be termed homotrophic". The term has not survived, although the notion underlying it is important, reflecting as it does pleiotropy, epistasis, embryonic inductive interactions, and constraint [Raff, 1996; Hall, 1983; 1999a; Burian *et al.*, 2000; Hall, 1999a; 2001a; Robert *et al.*, 2001; Hall and Olson, 2003].

later, Mivart [1871] had withdrawn his opposition, cited Lankester's homogeny and homoplasy favorably, and reduced his own classes of homology to three.[16]

7.1 Homoplasy vs Homology

Homoplasy is now contrasted with homology. "Homology and homoplasy are terms that travel together; homoplasy being close to, but not quite, the inverse of homology. If homology is "the same thing"... homoplasy is the *appearance* of "sameness" that results from independent evolution" [Wake, 1996, xvii].

The common basis for considering features as homoplastic is their independent evolution one from the other. However, homoplasy is a portmanteau term for classes of similarity otherwise subsumed under terms such as convergence, parallelisms, reversals, rudiments, vestiges and atavisms.[17] But these terms include features that may or may not be present (i) in all individuals within a population, all populations of a species, or all species within a genus, (ii) in all or any individuals of the most recent ancestor or more basal taxa, (iii) as fully functional structures or behaviours, or (iv) in taxa that share a common evolutionary history. This is as wide a set of categories as you will find anywhere in biology. Does homoplasy really encompass them all or is our understanding of homoplasy misplaced, incomplete or perhaps even wrong headed?

In the balance of this entry I outline in brief the classes of similarity that traditionally are grouped as homoplasy, these being convergence, parallelism, reversals, atavisms, rudiments and vestiges (see [Hall, 2003a] for a more in depth discussion and examples). I then align these classes of homoplasy with homology and demonstrate the developmental and phylogenetic continuities that underlie the apparently dichotomous categories homology and homoplasy.

8 CONVERGENCE

Convergence is the appearance of similar features in independent lineages. 'Independent' is usually taken to mean not having a shared recent ancestry, leaving both independent and recent ill-defined. Consequently, convergence is often defined operationally by an example, such as the evolution of insect, bird and bat wings [Shubin *et al.*, 1997; Hall, 2005a].

Because they are independently evolved, we expect convergent features *not* to share similar developmental pathways or mechanisms. Given the single tree

[16] Others proffered other terms to deal with the application of homology in particular fields of biology or for particular situations: *homogenesis* for the similarity of offspring to their parents [Mitchell, 1910], *homogenetic* or normal morphogeny [morphology], in contrast to convergent morphogeny or mere convergence [Willey, 1911], *homogenic* for genes with single alleles [Fisher, 1928] and *homogenetic* for the pairing in hybrids of chromosomes from one of the original ancestors [Waddington, 1939].

[17] See Patterson [1982; 1988], Wake [1991], McShea [1996] Sanderson and Hufford [1996], Meyer [1999] and Hall [1999a; 2002a; 2003a] for discussion of various types of homoplasy.

of life, however, taxa possessing convergent features share an evolutionary history, a history that includes conserved regulator (developmental) genes, gene networks/cascades and epigenetic processes such as cell aggregation, migration, differentiation and morphogenesis.

Darwin saw the difficulties of separating convergence from analogy. Determination of relationships between major groups of organisms and the search during the 19^{th} C for the group that gave rise to the chordates/vertebrates was complicated enormously by the inability to determine whether segmentation had evolved once (i.e., was homologous) or more than once and so was convergent. The same was (and is still) true for the origination of mesoderm(s).[18]

Arthur Willey, best known for his studies on amphioxus (*Branchiostoma*), and motivated by the plethora of theories of the origin of the vertebrates, devoted a book to convergence [Willey, 1911].[19] As with naming a new species of *Homo* rather than assigning your find to an already named taxon, Willey noted that there was "more joy amongst morphologists over one attempt at genealogy than over ninety and nine demonstrations of convergence" [1911, 53]. Willey did not confine himself to listing cases of convergence but proposed "that homoplasy does not cover all the cases which are included under convergence in the wider acceptance of the term" [and that] "all homoplasy is convergence, but all convergence is not homoplasy" [1911, xii, 11]; Willey saw that phylogenetic relations had to be established before convergence could be considered.

9 PARALLELISM

Parallelism, which has been described as a "gray zone between homology and convergence" [Gould, 2002, 1088], describes the presence (or formation) of a feature in related lineages but not in their most recent common ancestor and/or sister group If the feature were present in the most recent common ancestor it would be a homologue. If the feature was present in a more distant lineage (but not in the most recent common ancestor, its reappearance would be convergence (above). Loss and reappearance of a feature in a lineage is reversal (see below). Clearly knowledge of phylogenetic relationships is critical to our ability to separate homologous, convergent and parallel features. The expectation is that "if [character states] are parallelisms, then they should be developmentally and genetically the same" [Kellogg and Shaffer, 1993, 412].

Simpson [1961, 78] defined parallelism as "the development of similar characters separately in two or more lineages of common ancestry and on the basis of, or channeled by, characteristics of that ancestry". The phrase "characteristics of

[18]See Ospovat [1995, 111–112] for Darwin on convergence and analogy, and see Bowler [1996], Freeman and Martindale [2002] and Martindale *et al.* [2004] for the origin(s) of mesoderm(s).

[19]Willey's is perhaps the only book devoted to convergence, although Gregory [1951] lists well over 100 examples of convergence, and the treatises on variation and variability by Bateson [1894] and Yablokov [1974] include examples of convergence. See Bowler [1996] for Willey's views on parallelism and convergence.

that ancestry" we would now take as reflecting the presence of similar genetic and developmental mechanisms in different lineages and the response of those lineages to internal or external (mutational or environmental) influences to produce similar phenotypes. Simpson [1965] saw no fundamental difference between parallelism and convergence, because, he thought, mutations would be more likely to produce similar effects in closely than in more distantly related organisms.[20]

An example that separates parallelism and convergence is the study by Parra-Olea and Wake [2001] of body (trunk) elongation in salamanders within the tribe Bolitoglossini. Elongation can arise through the elongation of existing vertebrae, as in the genus *Lineatriton*, or by an increase in vertebral number, as in the genera *Batrachoseps* and *Oedipina*. Knowledge of phylogenetic relationships, emphasized as so important by Lankester, Willey and Simpson, enabled Parra-Olea and Wake to determine that these phylogenetic patterns of trunk elongation represent *parallelism* within the genus *Lineatriton* but *convergence* within the family; it is all in the depth of the phylogenetic relationships.

10 REVERSALS

To be identified as a reversal, rudiment, vestige or atavism a feature must bear a high degree of similarity to a character found in an ancestor,[21] a circumstance that suggests a conserved feature and so raises suspicions concerning relationship among reversals, rudiments, vestiges and atavisms, and between homology or homoplasy. Atavisms speak to us of reversibility of the phenotype associated with retention of developmental information. Rudiments, vestiges and atavisms also speak to retention of developmental information. Indeed, atavisms and reversals — and neomorphs, which are novel features found in a few individuals, but characteristic of all individuals in more derived taxa — can co-occur in the same population, as demonstrated elegantly for tarsal (ankle) bones in the salamander, *Taricha granulosa*, by Shubin *et al.* [1995].

A *reversal* is a reversion to a previous evolutionary state, a feature found in all the *adult* members of a taxon, in phylogenetically related lineages or in an ancestral lineage but not in the most recent common ancestor of those lineages. The designation reversal implies a phenotype in a descendent that develops from a developmental programme retained from an ancestor but not expressed in one or more intervening taxa. Phylogenetic character reversal is a synonymous name/process.[22] An oft-cited example, cited here again, is the loss of the second molar tooth in felids in the Miocene and its reappearance in the extant lynx, *Felis lynx* [Kurtén,

[20] See Kellogg and Shaffer [1993], Laporte [2000] and Hall [2003a] for discussion and amplification of Simpson's views.

[21] Of course, in many if not most cases, it may be difficult if not impossible to identify the common ancestor. In practice, a sister group or outgroup is used as the surrogate that indicates phylogenetic affinity.

[22] See McShea [1996], Bejder and Hall [2002] and Hall [2003a] for reversals, and see Stiassny [1992] for phylogenetic character reversals.

1963]. Reversals are regarded in the literature as homoplastic because the most recent ancestor or closely related extant taxa lack the feature, the implication being that the reversal is the result of independent evolution.

Dollo [1893], Meyrick [1927] and others claimed that reversal of evolution is not possible — hence Dollo's law [Hall, 2002b] — but they do not mean reversals in the sense outlined here. Dollo did not deny reversibility entirely, only that complex structures could not be recreated in their original form. But, if the developmental programme is retained, and/or if a vestige of the structure remains (as, for example, limb buds in snake embryos), the entire developmental programme does not have to re-evolve; it was never lost completely.[23]

11 RUDIMENTS AND VESTIGES

Although rudiment or vestige (vestigial feature/structure) are often used as synonyms, they can and should be distinguished on the basis of whether they occur in embryos or adults.

Rudiments are partly formed features that exist more fully (or fully) formed in ancestral or related taxa. Rudiments are found in embryos, are usually nonfunctional, but may serve a different function than did the ancestral feature. Examples include limb buds in snake embryos, hind limb buds in whale embryos, and tooth buds in toothless (baleen) whales. Atavisms arise from rudiments.

A *vestige* is a remnant in an adult of a feature that is more fully formed in ancestral or related taxa. Examples include the reduced eyes of blind cave dwelling fish and invertebrates, the reduced wings of flightless birds, and the reduced clavicles (collar bones) of mammals such as sheep and whales. The notochord, a vestige of the ancestral chordate dorsal skeleton that existed before the cartilaginous vertebral column evolved, is retained because of its major role in the induction of the embryonic dorsal nervous system. Other vestiges have different roles than the feature performed in the ancestral taxa. Some changes of role are subtle — the reduced wings of flightless birds are used in balance, display and courtship; penguins and cormorants use their wings to steer and propel themselves underwater.[24]

We are fortunate in having taxa in which the pattern (and perhaps the process) of vestigilization can be observed. Limb and/or digit reduction in lizards provides a particularly nice data set [Greer, 1987]. *Hemiergis peronii*, a lizard found in Western Australia, is polymorphic for digit number, individual lizards having two, three, four or five digits. Individuals with two or three digits also possess a vestigial fifth digit [Shapiro and Carl, 2001; Shapiro, 2002; Shapiro *et al.*, 2003].

[23]See Gregory [1936], Gould [1970], Hall [1984; 2002b; 2003a], Marshall *et al.* [1994], Lee and Shine [1998] and Teotónio and Rose [2001] for Dollo's Law and the reversibility of evolution.

[24]See Riedl [1978] and Hall [1983; 2005a] for the notochord as a vestige. For discussion and examples of rudiments and vestiges, see Berzin [1972], Yablokov [1974], Klima [1990], Hall [1984; 1999a; 2001b; 2002; 2003a] and Bejder and Hall [2002]. See Hall [2003a,b] for a discussion of the problems of character assessment when dealing with rudiments and vestiges, and Hall [2003a,b; 2005b] for the possibility that the feature in question may be a neomorph that mimics an ancestral feature.

12 ATAVISMS

Retention of developmental information is a means to maintain genetic diversity and retain the potential to introduce morphological change. An *atavism* is a feature that was present in an ancestral taxon and which reappears in low frequency in one or more individuals of a derived taxon; elements representing portion of the skeleton of the hind limbs seen in whale ancestors are found in 0.02 % of sperm whales [Berzin, 1972]. Three toes in modern horses, dew claws in dogs, and accessory nipples in mammals are others. An atavism or character reversal could arise in an organism with a rudiment of that feature if a chance mutation or an environmental cue that released hidden genetic variation, prolonged and expanded development of the rudiment, and if that new phenotype were selected for.[25]

13 THE CONTINUUM

Homology and homoplasy can be assigned at one level of the biological hierarchy, for example, the phenotype, without implying or prejudging statements about homology or homoplasy at other levels, e.g., the developmental or genetic basis of the feature. At the level of developmental processes, shared and divergent development can be two classes of homology because homologous features can arise from pathways that have diverged. Therefore we cannot distinguish homology from homoplasy on the basis that homologous features share a common development but homoplastic features do not. Parallelism, reversals, rudiments, vestiges and atavisms are features that form using similar developmental processes. Convergence is not based on shared development but reflects different processes producing similar features.

As outlined at the outset of this entry, we set homology against homoplasy in a dichotomy in which homoplasy includes (subsumes) convergence, parallelism, reversals, rudiments, vestiges and atavisms. Hall [2007a] suggested that examining nearness of relationships and degree of shared development reveals a continuum within the expanded homology category. The more refined version of that analysis set out by Hall [2003a] and as outlined in this entry, emphasizes a continuum from homology to parallelism, with convergence as the sole class of homoplasy.

The combined developmental/phylogenetic approach to homology and homoplasy outlined here resolves into: (1) Homology, which reflects phylogenetic conservation or retention of features in organisms with common descent and which subsumes parallelism, reversals, rudiments, vestiges and atavisms (i.e., all classes except convergence) as similar (homology), and (2) Convergence, which reflects

[25] See Struthers [1881], Evans [1955], Howell [1970], Raikow [1975], Riedl [1977], Lande [1978], Hall [1984; 1995b; 1999a; 2002a; 2003a], McKitrick [1986], Stiassny [1986; 1992], Wake and Larson [1987], Wyss [1988], Shubin *et al.* [1995], Tintant and Devillers [1995], Verhulst [1996], Greene and Cundall [2000] and Bejder and Hall [2002] for discussion and examples of atavisms. See Hallgrímsson and Hall [2005] and Hallgrímsson *et al.* [2005] for variation and the generation of features of the phenotype.

similar features resulting from independent evolution (homoplasy). Category 1, similarity due to common descent, is homogeny as proposed by Lankester. Category 2, similarity arising by independent evolution, is very close to homoplasy as proposed by Lankester to include analogy, parallelism and convergence. Remove parallelism from Lankester's homoplasy and you have homoplasy as I construe it.

This realignment of the categories of homologous and homoplastic features provides a way to bridge phylogenetic and developmental approaches to homology and homoplasy. Seeing reversals, rudiments, vestiges, atavisms and parallelism as closer to homology than to homoplasy should frame our agenda when searching for commonalities underlying these features. Regarded them as homoplastic sets our mind to independent evolution and directs us to search for different developmental and genetical mechanisms underlying homoplasy than underlie homology, an approach that is essentially pre-Darwinian and neglects the century and a half of evidence for a single evolutionary history of life on the planet. There is but one set of historically contingent phylogenetic and mechanistic relationships of features and taxa. Homology and homoplasy represent a continuum. Understanding that continuum is at the very foundation of empirical, historical and philosophical approaches to biology.

ACKNOWLEDGEMENTS

I came to these views on homology and homoplasy over the years following a workshop (held in Columbus. Ohio in 1999) on "Homoplasy in Primate and Human Evolution", organized by Charles Lockwood and John Fleagle and sponsored by the Wenner-Gren Foundation for Anthropological Research, and by the Institute of Human origins at Arizona State University. My thanks Charles and John for providing such a creative workshop environment. I thank Matt Vickaryous for comments on a draft manuscript and for his attempts (resisted by me more than he would like) to keep me on the cladistic straight and narrow. The research underlying this paper has been supported, most recently, by the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Killam Fellowship programme of the Canada Council for the Arts, and through funds associated with the University Research Professorship and Campbell Chair in Biology of Dalhousie University.

Box 1
Latent Homology

Latent homology [de Beer, 1971] is used when a feature in an ancestral taxon can be shown to have given rise to a structurally different but homologous feature in a descendant taxon. Rudiments that occasionally give rise to atavisms could be considered as latent homologues of those atavisms. The implication is that both features arise through developmental mechanisms that have been conserved but modified; the developmental processes have shown descent with modification. This was clear to Fritz Müller, author of *Für Darwin* [1864], the first published test of Darwin's theory of descent through modification of embryological development: "From the beginning of all things the Creator knew, that one day the inquisitive children of men would grope about after analogies and homologies, and that Christian naturalists would busy themselves with thinking out his Creative ideas; at any rate, in order to facilitate the discernment by the former that the opercular peduncle of the *Serpulae* is homologous with a branchial filament, He allowed it to make a *détour* in its development, and pass through the form of a barbate branchial filament" ([F. Müller, 1869, 114], English translation of [Müller, 1864]).

Mammalian middle ear ossicles are a classic example of latent homology, these ossicles being homologues of bones of the lower jaws of mammal-like reptiles (therapsids), a homology that would go unrecognized were it not for the fact that the transformation is documented in the fossil record and in the ontogeny of extant marsupials. Visceral arches in agnathan vertebrates/jaws in gnathostomes, and anterior appendages in early arthropods/ mouth parts in crustaceans are two other oft-cited examples. A more recently studied example is the search among uro- and cephalochordates for the precursors (latent homologues) of the vertebrate neural crest and neural crest cells.[a]

Henry Fairfield Osborn [1902] treated homoplasy as a law of latent homology, a view that only a few continue to recognize: "Homoplasy is an alternative perspective on homology, and when we can identify a phenomenon as latent homology we begin to approach an understanding of how homoplasy relates to homology on the one hand and to the production of diversity on the other" [D. B. Wake, 1999, 45]. In practice, in cladistic analyses, the ancestral and descendant features are identified as homoplastic because they are coded as independent or apomorphic characters. Coding the developmental processes and resulting features as multiple character states reveals the latent homologue.[b]

[a] See Gregory [1913], Hopson [1966], Thomson [1966], Lombard and Bolt [1979], Maier [1990] and Hall [2005a, especially Box 13.1] for the middle ear ossicles, Hall [1999a; 2002a] and Abzhanov and Kaufman [2000] for visceral arches and arthropod appendages, respectively, and Hall [1998; 1999b; 2000b; 2005b], Holland and Holland [2001] and Stone and Hall [2004] for the neural crest.

[b] See Hall [1995a; 2002c; 2003a; 2003b; 2005c] and Stone and Hall [2004, and references therein] for coding and latent homology.

Box 2
von Baer's Laws

von Baer's laws had their origin in a parallelism between human embryonic development and the history of life proposed independently by the German, J. F. Meckel [1811] and the Frenchman, E. R. A. Serres [1824; 1830]. Sometimes referred to as the Meckel's-Serres' law of parallelism, human embryonic development — and, as extended by Serres, to the development of non-mammalian vertebrates — progresses through a hierarchical series of animal types, each embryonic stage in turn representing fish, reptilian, mammalian, and finally human stages of human evolution. The progress of life on earth, which was being revealed during the 19^{th} C through discoveries of fossilized remains, paralleled this progression of embryonic development, not in any trivial way, but as a reflection of underlying common natural laws.[a]

von Baer [1828; 1835] went well beyond either Meckel or Serres to extend the concept of parallelism to the entire animal kingdom. Because some features in animals thought to be 'higher' up (more deeply nested in) the evolutionary tree were not present in animals considered 'lower' down the tree (more basal) — the avian yolk sac is not found in frog embryos, and birds were considered 'higher' up the evolutionary tree than frogs — von Baer concluded that higher forms do *not* recapitulate lower forms. Organs not organisms are recapitulated during embryonic development: "The embryo successively adds the organs that characterize the animal classes in the ascending scale. When the human embryo, for instance, is but a simple vesicle, it is an infusorian; when it has gained a liver, it is a mussel; with the appearance of the osseous system, it enters the class of fishes; and so forth, until it becomes a mammal and then a human being" ([Ospovat, 1976, 4–5], citing Oken's summary of von Baer's theory of recapitulation). Meckel and Serres independently proposed a law of comparison and recapitulation between levels of the natural world and proposed a universal mechanism of parallel change. von Baer's law was a statement about parallelism — development paralleled the classification of organisms — von Baer's theory, a theory of development (ontogeny) with evolutionary (phylogenetic) implications, not a law of recapitulation.

von Baer's law had several elements: (1) Embryonic stages are highly conserved. (2) Because development proceeds from the general to the specific,

earlier stages of embryos from different groups of organisms are more alike than are later stages. (3) As a consequence of 1 and 2, development parallels the classification of the same organisms into 'natural' groups. The first features to appear in development are those of the phylum, followed by features of the class, order, family, genus, and species; Matt Vickaryous, my cladistic conscience, would prefer that I write: "The first features to appear in development are those of the largest, most inclusive clade, followed by features of successively less inclusive clades".

The consequences of von Baer linking homology, development and classification were enormous, especially for the primacy of embryonic criteria in determining homology. E. S. Russell captured this beautifully: "And the surest way to determine the true homologies of parts will be to study their early development... Parts therefore, which develop from the same 'fundamental organ', and in the last resort from the same germ-layer, have a certain kinship, which may even reach the degree of exact homology" [Russell, 1916, 126]. The rapid dissemination of von Baer's laws into English by Martin Barry and William Carpenter, their adoption into the conceptualization of the archetype and of homology by Richard Owen, and into descent with modification by Charles Darwin, brought embryos and embryonic development to the forefront. Archetypes were to be sought in embryos, homology was to be sought in embryonic development, and the embryological criterion of homology was born.[b]

[a] See Russell [1916], Oppenheimer [1967], Gould [1977] and Bowler [1996] more for Meckel and Serres.

[b] See Desmond [1989], Hall [1992; 1999a], Bowler [1996], Richards [1987a,b] and Richmond [2000] for evaluations of how von Baer's laws were received by English morphologists.

BIBLIOGRAPHY

[Abouheif, 1997] E. Abouheif. Developmental genetics and homology: a hierarchical approach. *Trends Ecol. and Evol.*, 12: 405–408, 1997.

[Abouheif, 1999] E. Abouheif. Establishing homology criteria for regulatory gene networks: prospects and challenges. In G. R. Bock and G. Cardew (eds.), *Homology*, Novartis Foundation Symposium 222, pages 207–221. Wiley, Chichester, 1999.

[Abzhanov and Kaufman, 2000] A. Abzhanov and T. C. Kaufman. Homologs of *Drosophila* appendage genes in the patterning of arthropod limbs. *Devel. Biol.*, 227: 673–689, 2000.

[Amundson, 2007] R. Amundson (ed.). Reissue of *On the nature of Limbs* by R. Owen, with introductory essays and a Foreword by B. K. Hall. The University of Chicago Press, Chicago (in press), 2007.

[von Baer, 1828] K. E. von Baer. *Über Entwickelungsgeschichte der Thiere: Beobachtung und Reflexion*, Gebrüder, Bornträger, Könisberg. Reprinted 1967 by Culture et Civilisation, Bruxelles, 1828.

[von Baer, 1835] K. E. von Baer. *Untersuchungen über die Entwickelungsgeschichte der Fische.* Leipzig, 1835.

[Barry, 1836-1837a] M. Barry. On the unity of structure in the animal kingdom. *Edinburgh New Philos. J.*, 22: 116–141, 1836-1837.

[Barry, 1836-1837b] M. Barry. Further observations on the unity of structures in the animal kingdom, and on congenital anomalies, including 'Hermaphrodites'; with some remarks on

embryology, as facilitating animal nomenclature, classification, and the study of comparative anatomy. *Edinburgh New Philos. J.*, 22: 345–364, 1836-1837.

[Bateson, 1894] W. Bateson. *Materials for the study of variation, treated with especial regard to discontinuity in the origin of species*. Macmillan & Co., London. 598 pp., 1894. [Reprinted 1992 with an introduction by P. J. Bowler and an essay by G. Webster, Johns Hopkins University Press, Baltimore.]

[Bejder and Hall, 2002] L. Bejder and B. K. Hall. Limbs in whales and limblessness in other vertebrates: mechanisms of evolutionary and developmental transformation and loss. *Evol. & Devel.*, 4: 445–458, 2002.

[Berzin, 1972] A. A. Berzin. *The Sperm Whale*. Keter Press, Jerusalem, 1972.

[Bock and Cardew, 1999] G. R. Bock and G. Cardew (eds). *Homology*. Novartis Foundation Symposium 222. John Wiley & Sons, Chichester, 1999.

[Bock, 1991] W. J. Bock. Explanations in *Konstruktionsmorphologie* and evolutionary morphology. In N. Schmidt-Kittler and K. Vogel (eds.), *Constructional Morphology and Evolution*. Springer-Verlag, Berlin, pages 9–29, 1991.

[Bock and von Wahlert, 1965] W. J. Bock and G. von Wahlert. Adaptation and the form-function complex. *Evolution*, 19: 269–299, 1965.

[Bolker, 1992] J. A. Bolker. Constancy and variation in developmental mechanisms: an example from comparative embryology. In J. Mittenthal and A. Baskin (eds.), *Principles of Organization in Organisms*. Volume 13, pages 73–85. Addison-Wesley, Menlo Park, CA., 1992.

[Bowler, 1984] P. J. Bowler. *Evolution. The History of an Idea*. University of California Press, Berkeley, 1984.

[Bowler, 1996] P. J. Bowler. *Life's Splendid Drama. Evolutionary Biology and the Reconstruction of Life's Ancestry, 1860-1940*. The University of Chicago Press, Chicago, 1996.

[Boyden, 1943] A. Boyden. Homology and analogy: a century after the definitions of "homologue" and "analogue" of Richard Owen. *Quart. Rev. Biol.*, 18: 228–241, 1943.

[Boyden, 1947] A. Boyden. Homology and analogy. A critical review of the meanings and implications of these concepts in biology. *Amer. Midl. Nat.*, 37: 648–669, 1947.

[Burian *et al.*, 2000] R. M. Burian, S. F. Gilbert, P. M. Mabee, and B. J. Swalla (eds.). Evolutionary Developmental Biology: Paradigms, Problems and Prospects. *Amer. Zool.*, 40: 711–831, 2000.

[Butler and Saidel, 2000] A. B. Butler and W. M. Saidel. Defining sameness: historical, biological, and generative homology. *BioEssays*, 22: 846–853, 2000.

[Camardi, 2001] G. Camardi. Richard Owen, morphology and evolution. *J. Hist. Biol.*, 34: 481–515, 2001.

[Carpenter, 1839] W. B. Carpenter. *Principles of General and Comparative Physiology*. John Churchill, London, 1839.

[Carroll *et al.*, 2005] S. B. Carroll, J. K. Grenier, and S. D. Weatherbee. *From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design*. Second Edition. Blackwell Publishing, Malden, MA., 2005.

[Clack, 2002] J. A. Clack. *Gaining Ground. The Origin and Evolution of Tetrapods*. Indiana University Press, Bloomington and Indiana, 2002.

[Clark, 1900] J. W. Clark. *Old Friends at Cambridge and Elsewhere*. MacMillan and Co. Ltd., London, 1900.

[Cohen, 1994] I. B. Cohen. *Interactions. Some Contacts between the Natural Sciences and the Social Sciences*. The MIT Press, Cambridge, MA., 1994

[Cole and Hall, 2004a] A. G. Cole and B. K. Hall. *Cartilage is a metazoan tissue; integrating data from non-vertebrate sources. Acta Zool.*, 85: 69–80, 2004.

[Cole and Hall, 2004b] A. G. Cole and B. K. Hall.The nature and significance of invertebrate cartilages revisited: Distribution and histology of cartilage and cartilage-like tissues within the Metazoa. *Zoology*, 107: 261–274, 2004.

[Darwin, 1910] C. Darwin. *The Origin of Species by Means of Natural Selection*, John Murray London, 1910.

[Darwin and Seward, 1903] F. Darwin and A. C. Seward. *More Letters of Charles Darwin*, 2 Volumes, John Murray, London, 1903.

[de Beer, 1938] G. R. de Beer. Embryology and evolution. In G. R. de Beer (ed.), *Evolution: Essays on Aspects of Evolutionary Biology presented to Professor E. S. Goodrich on his Seventieth Birthday*, pages 59–78. The Clarendon Press, Oxford, 1938.

[de Beer, 1971] G. R. de Beer. *Homology: an Unsolved Problem*. Oxford Biology Reader No. 11, Oxford University Press, Oxford, 1971.

[Desmond, 1982] A. Desmond. *Archetypes and Ancestors. Palaeontology in Victorian London. 1850-1875*. The University of Chicago Press, Chicago, IL., 1982

[Di Gregorio, 1995] M. A. Di Gregorio. A wolf in sheep's clothing: Carl Gegenbaur, Ernst Haeckel, the vertebral theory of the skull, and the survival of Richard Owen. *J. Hist. Biol.*, 28: 247–280, 1995.

[Dollo, 1893] L. Dollo. Les lois de l'évolution. *Bull. Belgian Soc. for Geol. Palaeont. Hydrology*, 8: 164–166, 1893.

[Duboule and Wilkins, 1998] D. Duboule and A. S. Wilkins. The evolution of 'bricolage'. *Trends Genet.*, 14: 54–59, 1998.

[Endress *et al.*, 2000] P. K. Endress, P. Baas, and M. Gregory. Systematic plant morphology and anatomy—50 years of progress. *Taxon*, 49: 401–434, 2000.

[Evans, 1955] H. E. Evans. The osteology of a worm snake, *Typhlops jamaicensis* (Shaw). *Anat. Rec.*, 122: 381–396, 1955.

[Fisher, 1928] R. A. Fisher. The possible modification of the response of wild type to recurrent mutation. *Amer. Nat.*, 62: 115–126, 1928.

[Fitch, 2000] W. M. Fitch. Homology: a personal view on some of the problems. *Trends Genet.*, 16: 227–231, 2000.

[Freeman and Martindale, 2002] G. Freeman and M. Q. Martindale. Intracellular fate mapping and the origin of mesoderm in Phoronids. *Devel. Biol.*, 252: 301–311, 2002.

[Gans, 1985] C. Gans. Differences and similarities: comparative methods in mastication. *Amer. Zool.*, 25: 291–301, 1985.

[Gegenbaur, 1859] K. Gegenbaur. *Grundzüge der Vergleichenden Anatomie*, W. Engelmann, Leipzig, 1859. (2nd edition, 1970).

[Gegenbaur, 1870] K. Gegenbaur. *Grundzüge der Vergleichenden Anatomie*. Leipzig, 1870.

[Gegenbaur, 1878] K. Gegenbaur. *Elements of Comparative Anatomy*. (Translated by F. J. Bell), MacMillan Co., London, 1878.

[Gerhart, 2000] J. Gerhart. Inversion of the chordate body axis: are there alternatives? *Proc. Natl Acad. Sci. U.S.A.*, 97: 4445–4448, 2000.

[Ghiselin, 1976] M. T. Ghiselin. The nomenclature of correspondence: a new look at "homology" and 'analogy". In R. B. Masterton, W. Hodos, and H. Jerison (eds.), *Evolution, Brain and Behavior: Persistent Problems*, pages 129–142, Lawrence Erlbaum Assoc., Hillsdale, NJ., 1976.

[Gould, 1970] S. J. Gould, (). Dollo on Dollo's Law: Irreversibility and the status of evolutionary laws. *J. Hist. Biol.* **3**, 189-212, 1970.

[Gould, 1977] S. J. Gould. *Ontogeny and Phylogeny*. The Belknap Press of Harvard University Press, Cambridge, MA., 1977.

[Gould, 2002] S. J. Gould. *The Structure of Evolutionary Theory*. The Belknap Press of Harvard University Press, Cambridge MA., 2002.

[Greene and Cundall, 2000] H. W. Greene and D. Cundall. Limbless tetrapods and snakes with legs. *Science*, 287: 1939–1941, 2000.

[Greer, 1987] A. E. Greer. Limb reduction in the lizard genus *Lerista*. I. Variation in the number of phalanges and presacral vertebrae. *J. Herpetol.*, 21: 267–276, 1987.

[Gregory, 1913] W. K. Gregory. Critique of recent work on the morphology of the vertebrate skull, especially in relation to the origin of mammals. *J. Morphol.*, 24: 1–42, 1913.

[Gregory, 1936] W. K. Gregory. On the meanings and limits of irreversibility of evolution. *Amer, Natl.*, 70: 517–528, 1936.

[Gregory, 1951] W. K. Gregory. *Evolution Emerging. A Survey of ChangingPatterns from Primeval Life to Man*. Two Volumes. The Macmillan Company, New York, 1951.

[Guralnick, 2002] R. Guralnick. A recapitulation of the rise and fall of the cell lineage research program: The evolutionary-developmental relationship of cleavage to homology, body plans and life history. *J. Hist. Biol.*, 35: 537–567, 2002.

[Haas and Simpson, 1946] O. Haas and G. G. Simpson. Analysis of some phylogenetic terms, with attempts at redefinition. *Proc. Amer. Philos. Soc.*, 90: 319–347, 1946.

[Haeckel, 1866] E. Haeckel. *Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformite Descendenz Theorie*. 2 Volumes, Georg Reimer, Berlin, 1866.

[Haeckel, 1876] E. Haeckel. *The History of Creation: or, the development of the Earth and its Inhabitants by the Action of Natural Causes: a popular Exposition of the Doctrine of Evolution in General and that of Darwin, Goethe, and Lamarck in Particular*. 2 Volumes, Appleton, New York, 1876.

[Hall, 1983] B. K. Hall. Epigenetic control in development and evolution. In B. C. Goodwin, N. Holder, and C. C. Wylie (eds.), *Development and Evolution*, pages 353–379, Cambridge University Press, Cambridge, 1983.

[Hall, 1984] B. K. Hall. Developmental processes underlying heterochrony as an evolutionary mechanism. *Can. J. Zool.*, 62: 1–7, 1984.

[Hall, 1992] B. K. Hall. *Evolutionary Developmental Biology*, Chapman and Hall, London, 1992.

[Hall, 1994a] B. K. Hall. *Homology: The Hierarchical Basis of Comparative Biology*, Academic Press, San Diego, 1994.

[Hall, 1994b] B. K. Hall. Introduction. In B. K. Hall (ed.), *Homology: The Hierarchical Basis of Comparative Biology*. pages 1–19. Academic Press, San Diego, CA., 1994.

[Hall, 1995a] B. K. Hall. Homology and embryonic development. *Evolutionary Biology*, 28: 1–37, 1995.

[Hall, 1995b] B. K. Hall. Atavisms and atavistic mutations: evolutionary conservation of genetic and developmental information. *Nature Genet.*, 10: 126–127, 1995.

[Hall, 1998] B. K. Hall. Germ layers and germ-layer theory revisited. Primary and secondary germ layers, neural crest as a fourth germ layer, homology, and demise of the germ-layer theory. *Evol. Biol.*, 30: 121–186, 1998.

[Hall, 1999a] B. K. Hall. *Evolutionary Developmental Biology*. Second edition. Kluwer Academic Publishers, Dordrecht, Netherlands, 1999.

[Hall, 1999b] B. K. Hall. *The Neural Crest in Development and Evolution*. Springer-Verlag, New York, 1999.

[Hall, 2000a] B. K. Hall. Balfour, Garstang and de Beer: The first century of evolutionary morphology. *Amer. Zool.*, 40: 718–728, 2000.

[Hall, 2000b] B. K. Hall. The Evolution of the Neural Crest in Vertebrates. In L. Olsson and C.-O. Jacobson (eds.), *Regulatory Processes in Development*, Wenner-Gren International Series Volume 76, pages 101–113. The Portland Press, London, 2000.

[Hall, 2001a] B. K. Hall. The gene is not dead, merely orphaned and seeking a home. *Evol. & Devel.*, 3: 225–228, 2001.

[Hall, 2001b] B. K. Hall. Development of the clavicles in birds and mammals. *J. Exp. Zool.*, 289: 153–161, 2001.

[Hall, 2002a] B. K. Hall. Atavisms. In *Encyclopedia of Evolution*, (ed. in chief M. Pagel), Vol. 1, pages 86–87. Oxford University Press, New York, 2002.

[Hall, 2002b] B. K. Hall. Dollo's Law. In *Encyclopedia of Evolution*, (ed. in chief M. Pagel), Vol. 1, pages 287–288. Oxford University Press, New York, 2002.

[Hall, 2002c] B. K. Hall. Palaeontology and evolutionary developmental biology: A science of the 19th and 21st centuries. *Palaeontology*, 45: 647–669, 2002.

[Hall, 2003a] B. K. Hall. Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. *Biol. Rev. Camb. Philos. Soc.*, 78: 409–433, 2003.

[Hall, 2003b] B. K. Hall. The emergence of form: The shape of things to come. *Devel. Dynam.*, 228: 292–298, 2003.

[Hall, 2004] B. K. Hall. Evolution as the Control of Development by Ecology. In B. K. Hall, R. Pearson, and G. B. Müller (eds.) *Environment, Evolution and Development: Towards a Synthesis*, pages ix-xxiii. MIT Press, Cambridge MA., 2004.

[Hall, 2005a] B. K. Hall. *Bones and Cartilage: Developmental and Evolutionary Skeletal Biology*. Elsevier Academic Press, London, 2005.

[Hall, 2005b] B. K. Hall. A consideration of the neural crest and its skeletal derivatives in the context of novelty/innovations. *J. Exp. Zool. (Mol. Dev. Evol.)*, 304B: 548–557, 2005.

[Hall, 2005c] B. K. Hall. In Goethe's Wake: Marvalee Wake's Contribution to the Development and Evolution of a Science of Morphology. *Zoology* (special issue as *Festschrift* for Marvalee H. Wake), 108: 269–275, 2005.

[Hall, 2007a] B. K. Hall. Homoplasy and homology: Dichotomy or Continuum? In C. Lockwood and J. Fleagle (eds.), *Homoplasy in Primate and Human Evolution*. Cambridge University Press, Cambridge, 2007.

[Hall, 2007b] B. K. Hall. Tapping Many Sources: The Adventitious Roots of Evo-Devo in the 19^{th} Century. In M. D. Laubichler and J. Maienschein (eds.), *From Embryology to Evo-Devo. A History of Embryology in the 20th Century*, Symposium of the Dibner Institute for the History of Science, MIT. MIT Press, Cambridge, MA., 2007.

[Hall and Olson, 2003] B. K. Hall and W. M. Olson (eds.). *Keywords & Concepts in Evolutionary Developmental Biology*. Harvard University Press, Cambridge, MA., 2003.

[Hall *et al.*, 2004] B. K. Hall, R. Pearson, and G. B. Müller (eds.). *Environment, Development, and Evolution: Toward a Synthesis*. MIT Press, Cambridge MA., 2004.

[Hallgrímsson *et al.*, 2005] B. Hallgrímsson, J. J. Y. Brown, and B. K. Hall. The study of phenotypic variability: An emerging research agenda for understanding the developmental-genetic architecture underlying phenotypic variation. In: B. Hallgrimsson, and B. K. Hall (eds.), *Variation: A Central Concept in Biology*. Elsevier/Academic Press, New York, pages 525–551, 2005.

[Hallgrímsson and Hall, 2005] B. Hallgrímsson and B. K. Hall. *Variation: A Central Concept in Biology*. Elsevier/Academic Press, New York, pages 1–7, 2005.

[Holland and Holland, 2001] L. Z. Holland and N. D. Holland. Evolution of neural crest and placodes: amphioxus as a model for the ancestral vertebrate? *J. Anat.*, 199: 85–98, 2001.

[Hopson, 1966] J. A. Hopson. The origin of the mammalian middle ear. *Amer. Zool.*, 6: 437–450, 1966.

[Howell, 1970] A. B. Howell. *Aquatic Mammals. Their Adaptation to Life in the Water*. Dover Publications, New York, 1970. [A reprint of the 1930 edition.]

[Hubbs, 1944] C. T. Hubbs. Concepts of homology and analogy. *Amer. Natl,*, 78: 289–307, 1944.

[Huxley, 1864] T. H. Huxley. *Lectures on the Elements of Comparative Anatomy*. London, 1864.

[Huxley, 1894] T. H. Huxley. Owen's position in the history of Anatomical Science. In: *The Life of Richard Owen by his Grandson R. Owen*. Volume 2, pages 273–332, John Murray, London, 1894.

[Kellogg and Shaffer, 1993] E. A. Kellogg and H. B. Shaffer. Model organisms in evolutionary studies. *Syst. Biol.*, 42: 409–414, 1993.

[Klima, 1990] M. Klima. Rudiments of the clavicle in the embryos of whales (Cetacea). *Z. Säugetierkunde*, 55: 202–212, 1990.

[Kurtén, 1963] B. Kurtén. Return of a lost structure in the evolution of the felid dentition. *Soc. Sci. Fennica Comm. Biol.*, B 26: 1–12, 1963.

[Lande, 1978] R. Lande. Evolutionary mechanisms of limb loss in tetrapods. *Evolution*, 32: 73–92, 1978.

[Lankester, 1870a] E. R. Lankester. On the use of the term homology in modern zoology, and the distinction between homogenetic and homoplastic agreements. *Ann. Mag. Nat. Hist.*, 6, Ser. 4: 34–43, 1870.

[Lankester, 1870b] E. R. Lankester. On the use of the term "homology". *Ann. Mag. Nat. Hist.*, 6, Ser. 4: 342, 1870.

[Lankester, 1911] E. R. Lankester. Zoology. In *The Encyclopædia Britannica*, 11th ed., Vol. 28, p. 1029. The Encyclopædia Britannica Company, New York, 1911.

[Laporte, 2000] L. F. Laporte. George Gaylord Simpson. Paleontologist and Evolutionist. Columbia University Press, New York, 2000.

[Lee and Shine, 1998] M. S. Y. Lee and R. Shine. Reptilian viviparity and Dollo's Law. *Evolution*, 52: 1441–1450, 1998.

[Lombard and Bolt, 1979] R. E. Lombard and J. R. Bolt. Evolution of the tetrapod ear: An analysis and reinterpretation. *Biol. J Linn. Soc*, 11: 19–76, 1979.

[Maclise, 1846] J. Maclise. On the nomenclature of anatomy. *Lancet*, March 14: 298–301, 1846.

[Maier, 1990] W. Maier. Phylogeny and ontogeny of mammalian middle ear structures. *Neth. J. Zool.*, 40: 55–74, 1990.

[Marshall *et al.*, 1994] C. R. Marshall, E. C. Raff, and R. A. Raff. Dollo's Law and the death and resurrection of genes. *Proc. Natl. Acad. Sci. U.S.A.*,91: 12283–12287, 1994.

[Martindale *et al.*, 2004] M.Q. Martindale, K. Pang, and J.R. Finnerty. Investigating the origins of triploblasty: "Mesodermal" gene expression in a diploblastic animal, the sea anemone, *Nematostella vectensis* (phylum, Cnidaria; Class Anthozoa). *Development*, 131: 2463–2474, 2004.

[Matthen, 2000] M. Matthen. What is a hand? What is a mind? *Rev. Int. Philos.*, 214: 653–672, 2000.

[Mayr, 1982] E. Mayr. *The Growth of Biological Thought. Diversity, Evolution and Inheritance*. The Belknap Press of Harvard University Press, Cambridge MA., 1982.

[McGhee, 2000] J. D. McGhee. Homologous tails? Or tails of homology? *BioEssays*, 22: 781–785, 2000.

[McKitrick, 1986] M. C. McKitrick. Individual variation in the flexor cruris lateralis muscle of the Tyrannidae (Aves, Passeriformes) and its possible significance. *J. Zool*, 209: 251–270, 1986.

[McShea, 1996] D. W. McShea. Complexity and homoplasy. In M. J. Sanderson and L. Hufford (eds.), *Homoplasy: the Recurrence of Similarity in Evolution*, pages 207–225. Academic Press, San Diego, 1996.

[McShea, 2000] D. W. McShea. Functional complexity in organisms: parts as proxies. *Biol. & Philos.*, 15: 641–668, 2000.

[Meckel, 1811] J. F. Meckel. *Entwurf einer Darstellung der zwischen dem Embryozustande der höheren Tiere und dem permanenten der niederen stattfindenen Parallele:Beyträge zur vergleichenden Anatomie*. Volume 2. Karl Heinrich Reclam, Leipzig, 1811

[Meyer, 1999] A. Meyer. Homology and homoplasy: The retention of genetic programmes. In G. R. Bock and G. Cardew (eds.), *Homology*, Novartis Foundation Symposium 222, pages 141–157. Wiley, Chichester, 1999.

[Meyrick, 1927] E. Meyrick. *A Revised Handbook of British Lepidoptera*. London, 1927.

[Minsuk and Keller,, 1996] S. B. Minsuk and R. E. Keller. Dorsal mesoderm has a dual origin and forms by a novel mechanism in *Hymenochirus*, a relative of *Xenopus*. *Devel. Biol.*, 174: 92–103, 1996.

[Mitchell, 1910] P. C. Mitchell. Biogenesis. In *The Encyclopædia Britannica*, 11th ed, vol. 3, page 952. The Encyclopædia Britannica Company, New York, 1910.

[Mivart, 1870] St. G. Mivart. On the use of the term "homology". *Ann. Mag. Nat. Hist.*, ser. 4, 6: 112–127, 1870.

[Mivart, 1871] St. G. Mivart. *On the Genesis of Species*. Appleton and Co., New York, 1871.

[Moment, 1945] G. B. Moment. The relationship between serial and special homology and organic similarities. *Amer. Natl.*, 79: 445–455, 1945.

[Moore, 1986] J. A. Moore. Science as a way of knowing-Developmental Biology. *Amer. Zool.*, 27: 1–159, 1986.

[Müller, 1864] F. Müller. *Für Darwin*. Engelmann, Leipzig, 1864.

[Müller, 1869] F. (J. F. T) Müller. *Facts and Arguments for Darwin*. Translated from the German by W. S. Dallas. John Murray, London, 1869.

[Newman and Müller, 2000] S. A. Newman and G. B. Müller. Epigenetic mechanisms of character origination. *J. Exp. Zool. (Mol. Dev. Evol.*, 288: 304–317, 2000.

[Nyhart, 1995] L. K. Nyhart. *Biology Takes Form. Animal Morphology and the German Universities, 1800-1900*. The University of Chicago Press, Chicago, 1995.

[Oppenheimer, 1959] J. Oppenheimer. An embryological enigma in the *Origin of Species*. In B. Glass, O. Temkin, and W. L. Strauss Jr. (eds.), *Forerunners of Darwin, 1745-1859*, pages 292–322, Johns Hopkins University Press, Baltimore, 1959.

[Oppenheimer, 1967] J. Oppenheimer. *Essays in the History of Embryology and Biology*, MIT Press, Cambridge, MA., 1967

[Osborn, 1902] H. F. Osborn. Homoplasy as a law of latent or potential homology. *Amer. Natl.*, 36: 259–271, 1902.

[Ospovat, 1976] D. Ospovat. The influence of Karl Ernst von Baer's Embryology, 1825-1859: A Reappraisal in Light of Richard Owen's and William B. Carpenter's 'Palaeontological Application of "Von Baer's Law"'. *J. Hist. Biol.*, 9: 1–28, 1976.

[Ospovat, 1995] D. Ospovat. *The Development of Darwin's Theory. Natural History, Natural Theology, and Natural Selection, 1838-1859*. Cambridge University Press, Cambridge, 1995.

[Owen, 1843] R. Owen. *Lectures on Comparative Anatomy and Physiology of the Invertebrate Animals, delivered at the Royal College of Surgeons in 1843*, Longman, Brown Green and Longman, London, 1843.

[Owen, 1846] R. Owen. Report on the Archetype and Homologies of the Vertebrate Skeleton. Report of thse sixteenth meeting of the *British Association for the Advancement of Science, pages 169–340*, John Murray, London, 1846.

[Owen, 1848] R. Owen. *On the Archetype and Homologies of the Vertebrate Skeleton*. Richard and John E. Taylor, London, 1848.

[Owen, 1949] R. Owen. textitOn the Nature of limbs. A Discourse delivered on Friday February 9, at an evening meeting of the Royal Institution of Great Britain. John Van Voorst, London, 1949.

[Padian, 1997] K. Padian. The rehabilitation of Sir Richard Owen. *BioScience*, 47: 446–453, 1997.

[Panchen, 1992] A. L. Panchen. *Classification, Evolution, and the Nature of Biology*. Cambridge University Press, Cambridge, 1992.

[Panchen, 1999] A. L. Panchen. Homology — history of a concept. In G. R. Bock and G. Cardew (eds.), *Homology*. Novartis Foundation Symposium 222, pages 141–157. Wiley, Chichester, 1999.

[Parra-Olea and Wake, 2001] G. Parra-Olea and D. B. Wake. Extreme morphological ecological homoplasy in tropical salamanders. *Proc. Natl. Acad. Sci. U.S.A.*, 98: 7888–7891, 2001.

[Patterson, 1982] C. Patterson. Morphological characters and homology. In K. A. Joysey and A. E. Friday (eds.), *Problems of Phylogenetic Reconstruction*, pages 21–74. Academic Press, London, 1982.

[Patterson, 1988] C. Patterson. Homology in classical and molecular biology. *Molec. Biol. Evol.*, 5: 603–625, 1988.

[Raff, 1996] R. A. Raff. *The Shape of Life. Genes, Development, and the Evolution of Animal Form*. The University of Chicago Press, Chicago, 1996.

[Raikow, 1975] R. J. Raikow. The evolutionary reappearance of ancestral muscles as developmental anomalies in two species of birds. *Condor*, 77: 514–517, 1975.

[Rathke, 1839] H. Rathke. *Bemerkungen über die Entwickelung des Schädels der Wirbelthiere*, Königsberg, 1839.

[Reilly and Lauder, 1988] S. M. Reilly and G. V. Lauder. Atavisms and the homology of hyobranchial elements in lower vertebrates. *J. Morphol*, 195: 237–246, 1988.

[Richards, 1987a] E. Richards. A question of property rights: Richard Owen's evolutionism reassessed. *Brit. J. Hist. Sci.*, 20: 129–171, 1987.

[Richards, 1987b] R. J. Richards. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago University Press, Chicago, 1987.

[Richmond, 2000] M. L. Richmond. T. H. Huxley's criticism of German cell theory: An epigenetic and physiological interpretation of cell structure. *J. Hist. Biol.*, 33: 247–289, 2000.

[Riedl, 1978] R. Riedl. *Order in LIving Organisms. A Systems Analysis of Evolution*. Translated by R. P. S. Jefferies. John Wiley & Sons, Chichester, 1978.

[Rieppel, 1988] O. Rieppel. *Fundamentals of Comparative Biology*. Birkhäuser Verlag, Basel, 1988.

[Rieppel and Kearney, 2002] O. Rieppel and M. Kearney. Similarity. *Biol. J. Linn. Soc.*, 175: 59–82, 2002.

[Robert *et al.*, 2001] J. S. Robert, B. K. Hall, and W. M. Olson. Bridging the gap between developmental systems theory and evolutionary developmental biology. *BioEssays*, 23: 954–962, 2001.

[Rupke, 1993] N. Rupke. Richard Owen's vertebrate archetype. *Isis*, 84: 231–251, 1993.

[Rupke, 1994] N. A. Rupke. *Richard Owen: Victorian Naturalist*. Yale University Press, New Haven, CT., 1994.

[Russell, 1916] E. S. Russell. *Form and Function. A Contribution to the History of Animal Morphology*, John Murray, London, 1916. (Reprinted 1972 by Gregg International Publishers Ltd, Westmead, England, and in 1982 by The University of Chicago Press, Chicago, with a new introduction by George V. Lauder.)

[Sanderson and Hufford, 1996] M. J. Sanderson and L. Hufford (eds.). *Homoplasy: Recurrence of Similarity in Evolution*. Academic Press, San Diego, 1996.

[Sattler, 1984] R. Sattler. Homology – a continuing challenge. *Syst. Bot.*, 9: 382–394, 1984.

[Sattler, 1992] R. Sattler. Process morphology: structural dynamics in development and evolution. *Can. J. Bot.*, 70: 708–714, 1992.

[Sattler, 1994] R. Sattler. Homology, Homeosis, and process morphology in plants. In B. K. Hall (ed.), *Homology: The Hierarchical Basis of Comparative Biology*, pages 424–475. Academic Press, San Diego, 1994.

[Sattler and Jeune, 1992] R. Sattler and B. Jeune. Multivariate analysis confirms the continuum view of plant form. *Ann. Bot.*, 69: 249–262, 1992.

[Serres, 1824] E. R. A. Serres. Explication du système nerveus des animaux invertébrés. *Ann. Sci. Nat.* **3**, 377-380, 1824.

[Serres, 1830] E. R. A. Serres. Anatomie transcendante – Quartrième mêmoire: Loi de symétrie et de conjugaison du système sanguin. *Ann. Sci. Nat.*, 21: 5–49, 1830.
[Shapiro, 2002] M. D. Shapiro. Developmental morphology of limb reduction in *Hemiergis* (Squamata: Scincidae): Chondrogenesis, osteogenesis, and heterochrony. *J. Morphol.*, 254: 211–231, 2002.
[Shapiro and Carl, 2001] M. D. Shapiro and T. F. Carl. Novel features of tetrapod limb development in two nontraditional model species: A skink and a direct-developing frog. In M. L. Zelditch (ed.), *Beyond Heterochrony. The Evolution of Development*, pages 337–361. Wiley-Liss, New York, 2001.
[Shapiro *et al.*, 2003] M. D. Shapiro, J. Hanken, and N. Rosenthal. Developmental basis of evolutionary digit loss in the Australian lizard *Hemiergis*. *J. Exp. Zool. (Mol. Dev. Evol.)*, 297B: 48–56, 2003.
[Shubin *et al.*, 1995] N. Shubin, D. B. Wake, and A. J. Crawford. Morphological variation in the limbs of *Taricha granulosa* (Caudata: Salamandridae): Evolutionary and phylogenetic implications. *Evolution*, 49: 874–884, 1995.
[Shubin *et al.*, 1997] N. Shubin, C. Tabin, and S. Carroll. Fossils, genes, and the evolution of animal limbs. *Nature*, 388: 639–648, 1997.
[Simpson, 1961] G. G. Simpson. *Principles of Animal Taxonomy*. Columbia University Press, New York, 1961.
[Simpson, 1965] G. G. Simpson. *The Meaning of Evolution. A Study of the History of Life and of its Significance to Man*. Yale University Press, New Haven and London, 1965.
[Singer, 1959] C. Singer. *A History of Biology to about the year 1900. A General Introduction to the Study of Living Things*. 3rd and revised edition, Abelard-Schuman, London and New York, 1959.
[Smith, 1967] H. M. Smith. Biological similarities and homologies. *Syst. Zool.*, 16: 101–102, 1967.
[Spemann, 1915] H. Spemann. Zur Geschichte und Kritik des Begriffs der Homologie. Die Kultur der Gegenwar Teil III, Abt. IV. *Allgemeine Biol.*, 1: 63–86, 1915.
[Stevens, 1984] P. S. Stevens. Homology and phylogeny: morphology and systematics. *Syst. Bot.*, 9: 395–409, 1984.
[Stiassny, 1986] M. L. J. Stiassny. The limits and relationships of the acanthomorph teleosts. *J. Zool. Soc. Lond. (B)*, 1: 411–460, 1986.
[Stiassny, 1992] M. L. J. Stiassny. Atavisms, phylogenetic character reversals, and the origin of evolutionary novelties. *Neth. J. Zool.*, 42: 260–276, 1992.
[Stone and Hall, 2004] J. R. Stone and B. K. Hall. Latent homologues for the neural crest as an evolutionary novelty. *Evol. & Devel*, 6: 123–129, 2004.
[Striedter and Northcutt, 1991] G. F. Striedter and R. G. Northcutt. Biological hierarchies and the concept of homology. *Brain Behav. Evol.*, 38: 177–189, 1991.
[Struthers, 1881] J. Struthers. On the bones, articulations, and muscles of the rudimentary hind-limb of the Greenland Right-whale (*Balaena mysticetus*). *J. Anat. Physiol.*, 15: 141–176, 302–321, 1881.
[Svensson, 2004] M. E. Svensson. Homology and homocracy revisited: gene expression patterns and hypotheses of homology. *Dev. Genes Evol.*, 214: 418–421, 2004.
[Szarski, 1949] H. Szarski. The concept of homology in the light of the comparative anatomy of vertebrates. *Quart. Rev. Biol.*, 24: 124–131, 1949.
[Teotónio and Rose, 2001] H. Teotónio and M. R. Rose. Reverse evolution. *Evolution*, 55: 653–660, 2001.
[Thomson, 1966] K. S. Thomson. The evolution of the tetrapod middle ear in the rhipidistian-tetrapod transition. *Amer. Zool.*, 6: 379–397, 1966.
[Tintant and Devillers, 1995] H. Tintant and C. Devillers. Atavisms in present and past — its function in evolution. *Bull. Zool. France Evol. Zool.*, 120: 327–334, 1995.
[Valentine, 2004] J. W. Valentine. *On the Origin of Phyla*. The University of Chicago Press, Chicago and London, 2004.
[Verhukst, 1996] J. Verhukst. Atavisms in *Homo sapiens* — a Bolkian heterodoxy revisited. *Acta Biotheor,*, 44: 59–73, 1996.
[Waddington, 1939] C. H. Waddington. *An Introduction to Modern Genetics*. Allen and Unwin, London, 1939.
[Wagner, 2001] G. P. Wagner (ed.). *The Character Concept in Evolutionary Biology*. Academic Press, San Diego, 2001.

[Wagner, 2000] P. J. Wagner. Exhaustion of morphologic character states among fossil taxa. *Evolution*, 54: 365–386, 2000.
[Wake, 1991] D. B. Wake. Homoplasy: The result of natural selection or evidence of design limitations? *Amer. Natl.*, 138: 543–569, 1991.
[Wake, 1994] D. B. Wake. Comparative similarity. *Science*, 265: 268–269, 1994.
[Wake, 1996] D. B. Wake. Introduction. In M. J. Sanderson and L. Hufford (eds.), *Homoplasy: The Recurrence of Similarity in Evolution*, pages xvii-xxv. Academic Press, San Diego, CA., 1996.
[Wake, 1999] D. B. Wake. Homoplasy, homology and the problem of 'sameness' in biology. In G. R. Bock and G. Cardew (ed.), *Homology*. Novartis Foundation Symposium 222, pages 24–33. Wiley, Chichester, 1999.
[Wake and Larson, 1987] D. B. Wake and A. Larson. Multidimensional analysis of an evolving lineage. *Science*, 238: 42–48, 1987.
[Wake, 1996] M. H. Wake. The use of unconventional morphological characters in the analysis of systematic patterns and evolutionary processes. In L. Grande and O. Rieppel (eds.), *Interpreting the Hierarchy of Nature: From Systematic Patterns to Evolutionary Process Theories*, pages 173–200. Academic Press, San Diego, 1996.
[Wake, 2001] M. H. Wake. Bodies and body plans, and how they came to be. In W. J. Kress and G. W. Barrett (eds.), *A New Century of Biology*, pages 28–52. Smithsonian Institution Press, Washington and London, 2001.
[Wake and Summers, 2003] M. H. Wake and A. P. Summers. Morphology. In B. K. Hall and W. M. Olson (eds.) *Keywords & Concepts in Evolutionary Developmental Biology*, pages 268–275. Harvard University Press, Cambridge, MA., 2003.
[Weiss, 1994/1995] K. M. Weiss. What recapitulates what? Genetics and the evolution of development. *Evol. Anthropol.*, 3: 216–222, 1994/1995.
[Weiss and Buchanan, 2004] K. M. Weiss and A. V. Buchanan. *Genetics and the Logic of Evolution*. John Wiley & Sons, Hoboken, NJ., 2004.
[Wilkins, 2002] A. S. Wilkins. *The Evolution of Developmental Pathways*. Sinauer Associates, Sunderland, MA., 2002.
[Willey, 1911] A. Willey. *Convergence in Evolution*. John Murray, London, 1911.
[Wyss, 1988] A. R. Wyss. On "retrogression" in the evolution of the Phocinae and phylogenetic affinities of the Monk seals. *Amer. Mus. Novit.*, 2924: 1–38, 1988.
[Yablokov, 1974] A. V. Yablokov. *Variability of Mammals*. Amerind Publishing Co., New Delhi, 1974.
[Young, 1993] B. A. Young. On the necessity of an Archetypal concept in morphology: With special reference to the concepts of "Structure" and "Homology". *Biol. & Philos.*, 8: 225–248, 1993.

This page intentionally left blank

BIOLOGICAL CONCEPTIONS OF RACE

Robin O. Andreasen

1 INTRODUCTION

Do human races exist or are human racial categories biologically meaningless?[1] Questions such as this one are part of a series of more general philosophical questions such as those that concern realism and classification in the natural sciences. What is the basis of an objective classification scheme within systematic biology? In what ways, if any, do systematists's conceptions of biological kinds deviate from traditional philosophical conceptions of natural kinds? To what extent, if any, is everyday language about biological kind terms characterized by the practice of semantic deference to science? To what extent should it be?

Questions about the biological reality, or lack thereof, of race are also thought to carry important social implications. Part of the reason is that belief in the biological reality of race has been (and sometimes still is) used in an effort to justify sociopolitical inequalities. Though the connection between biological realism and racism is a contingent one, it is generally assumed that the nonexistence of biological races is an important first step in arguments against racism.

There are at least two ways to approach questions about the biological reality, or lack thereof, of race. One way is to examine conceptions of race within systematic biology and to ask whether any can be applied, successfully, to humans. Another way is to examine common sense conceptions of race and to ask whether biology provides support for any such conceptions.[2] For the most part, this chapter takes the first approach. Yet it is important to note that these two approaches are not entirely distinct. Throughout the history of the term, common sense and biological conceptions of 'race' have developed side by side — each reinforcing and influencing the other [Banton and Harwood, 1975; Smedley, 1993; Gould, 1981; Appiah, 1996]. In addition it is generally agreed that scientific conceptions of race must overlap reasonably with common sense conceptions; otherwise, we are no longer talking about 'race' [Zack, 2002; Glasgow, 2003].

Historically, few scholars doubted the biological reality of race. There have been at least two prominent race concepts in systematic biology — namely, the

[1] In this paper, I will use the expressions 'biologically objective', 'biologically real', and 'biologically significant' interchangeably. Though there are important differences in the meanings of these expressions, such differences will not be important for our purpose.

[2] For critical discussions of this type of approach see Mallon [forthcoming] and Haslanger [2005]. Both reject this semantic strategy in favor of pragmatic and normative approaches to understanding 'race'.

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

typological and the geographical concepts. Both have a long history and both have been used for defining human and nonhuman race. Today, however, the majority view is that human biological races don't exist. Nevertheless, a growing number of biologists, anthropologists, and philosophers accept the shortcomings of earlier biological definitions, but nonetheless maintain that human biological races might still exist.

2 THE TYPOLOGICAL RACE CONCEPT

The typological race concept represents one of the first comprehensive efforts to provide a biologically objective definition of 'race'. This concept has its roots in the philosophical doctrine of essentialism as well as pre-Darwinian ideas about the objective basis of systematic classification schemes. Essentialism is the idea that natural kinds ought to be individuated in terms of *kind-specific essences*. A 'kind-specific essence' can be roughly defined as a nonaccidental and intrinsic property (or set of such properties) that an object must have in order to be the kind of thing that it is. Such properties are supposed to be necessary and sufficient for kind-membership; they are also supposed to account for other properties typical of members of a kind. Take gold, for example. The essence of gold can be given by its atomic number. Being made of atoms that have atomic number 79 is a nonaccidental and intrinsic property possessed by all and only gold things. In addition, this property explains many other properties characteristic of gold such as its color and malleability.

Prior to the modern synthesis, it was generally agreed that essentialism is the right view to take about systematic classification [Hull, 1965; Mayr, 1942; 1959; Ereshefsky, 2001]. Naturalists used essentialism to define taxa at all levels in the taxonomic hierarchy. Its application to 'race' was no exception. As with any kind-specific essence, racial essences are supposed to be nonaccidental and intrinsic properties, possessed by all and only the members of a race, that account for many other properties typical of that race. In the case of humans, the essence of each race was assumed to be in the blood, melanin, or cranial shape and size [Banton and Harwood, 1975; Gould, 1981; Smedley, 1993].[3] A contemporary essentialist might argue for genetic essences. Many essentialists also assumed that the members of a race are similar not only with respect to overt physical features (skin color, hair type, eye shape, etc.), but with respect to psychological and behavioral traits as well [Banton and Harwood, 1975; Gould, 1981; Appiah, 1996]. It is important to note, however, that racial typology would not be shown false if psychological and behavioral traits were found to be poor predictors of race membership.

The common sense corollary to the typological race concept is what Anthony Appiah [1996] has called 'racialism'. Racialism is the idea that humans can be

[3]Although racial typologists assumed the existence of race-specific essences, they were hard-pressed to find any property or set of properties that could meet all of the criteria demanded of essential properties. Defenders of typological conceptions of species faced a similar problem [Ereshefsky, 2001].

divided into a small number of racial groups such that all of the members of each share certain heritable characteristics (skin color, etc.) with one another that they do not share with the members of any other race. The main difference between racialism and racial typology is that, unlike racial typology, racialism *would* be shown false if psychological and behavioral traits were found to be poor predictors of race membership.

Throughout the late nineteenth and early twentieth centuries, racial typology and racialism developed side by side, each influencing and reinforcing the other [Banton and Harwood, 1975; Gould, 1981; Smedley, 1993; Appiah, 1996]. For example, in a set of instructions written for explorers on how to study indigenous peoples, George Cuvier assumed that the physical differences among different groups would explain their social and behavioral differences. Later he elaborated this idea by arguing that humans are naturally divided into three racial types, which can be arranged in an ascending scale from blacks to Asians to whites [Cuvier, 1812]. Similar ideas were held by other nineteenth century race scholars. S. G. Morton [1839; 1844; 1849], for example, argued for statistically significant differences in the cranial capacities among members of different races. He also maintained that these data establish a biologically based racial hierarchy. J. C. Nott and G. R. Gliddon advanced a similar argument in a book titled *Types of Mankind* [1854] as did Count Joseph Arthur Gobineau [1853-54] in his *Essay on the Inequality of Human Races*.[4]

The typological race concept was the received biological race concept for more than a century. Today, most scholars agree that it is mistaken. Yet, there is some confusion over the exact nature of the problem. A number of biological arguments have been offered to explain where the typological concept goes wrong. Some of these are better than others. In the remainder of this section, I will critically examine such arguments with the aim of isolating the better ones.

Some race scholars have adapted an argument that was originally developed against the typological species concept (i.e., the idea that species taxa ought to be defined in terms of species-specific essences) and have applied it to the typological race concept.[5] Defenders of this argument maintain that racial typology is incompatible with contemporary evolutionary theory because the former supposes that races are static and unchanging while the latter supposes that races evolve [Montagu, 1941; Banton and Harwood, 1975; Goldberg, 1993; Smedley, 1993; Zack, 2002].[6] The assumption at work here is that biological races, if they exist, are taxonomic categories. Hence they, like species, are supposed to be capable of evolving.

[4]See Banton and Harwood [1975], Smedley [1993], Gould [1981], and Appiah [1996] for more on the history of racial typology. See Mayr [1959], Hull [1965], and Sober [1980] for more on the influence of typological thinking in other areas of systematic biology.

[5]See, for example, Hull [1965].

[6]When discussing arguments against the biological reality of race, some scholars use double quotations around the term 'race' to indicate their belief that races are biologically unreal. In this chapter I will not follow this convention in part because many scholars who reject the biological reality of race nonetheless maintain that races are socially real (See, for example, [Sundstrom, 2002] and [Root, 2000].

Racial typology, on the other hand, assumes that races are natural kinds defined according to essentialist criteria. Since such kinds are supposed to be immutable, defenders of this argument conclude that races cannot be types.

Although it is true that most nineteenth century typologists believed in the fixity of types, this assumption was not held by all typologists [Sober, 1980]. Both Aristotle and Linnaeus considered the possibility that a new species could arise as a result of cross-species hybridization. Likewise Cuvier, Smith, and Gobineau each defended the idea that all humans are the descendants of Adam and Eve and that races did not come into existence until after Adam's and Eve's initial creation [Banton and Harwood, 1975].[7] Perhaps more importantly, there is no inconsistency in the idea that essentialism is compatible with change [Sober, 1980]. According to essentialism, natural kinds are eternal categories that individual objects occupy. While each natural kind has its own immutable essence, an individual can change with respect to its accidental properties and still be a member of the same natural kind. An individual can also change with respect to its kind-specific essence — but when it does, it ceases to be a member of that natural kind. Applying these ideas to race, a population belonging to one race can give rise to a population belonging to a different race — but when this occurs, the two populations are members of different natural kinds.

A second argument advanced against the typological race concept is the idea that races cannot be types because the boundaries among races are vague [Banton and Harwood, 1975; Zack, 2002; Appiah, 1996]. Defenders of this argument maintain that many of the phenotypic traits used to individuate races (e.g., skin color, hair type, bone structure, etc.) are clinal; they vary gradually across a geographic range. Clinal variation is assumed to pose a problem for racial typology because natural kinds are supposed to have sharp boundaries; they are not supposed to shade into one another.[8]

As a number of philosophers have argued, however, the requirement that there be precise boundaries between natural kinds is no longer viewed as tenable [Boyd, 1991; Sober, 1993; Pigliucci and Kaplan, 2003]. It is like demanding that there be a precise line of demarcation between baldness and having a full head of hair, or between being rich and being poor. That there are line drawing problems in these cases does not mean that the properties in question (wealth and baldness) are unreal. Likewise natural kinds defined according to essentialist criteria might be real, even if the boundaries among them are somewhat vague.

Some scholars have tried to discredit racial typology by appeal to Mendel's law of independent assortment [Montagu, 1941; Banton and Harwood, 1975; Appiah, 1996; Zack, 1997; 2002]. Independent assortment is the idea that genes are assorted independently in meiosis, unless they are tightly linked on the same chromosome. Defenders of this argument assume that racial essences, if they exist, are genetic properties. They also assume that races must breed true to type. That is, the members of a race must share a number of traits with one another that they do

[7] Racial differentiation was assumed to be the result of a series of natural catastrophes.

[8] For a parallel argument against species essentialism see [Hull, 1965].

not share with the members of any other race — and this collection of traits is supposed to be passed on faithfully in reproduction. Because many of the traits typically used to individuate races (skin color, bone structure, etc.) are multigenic and are most likely spread all over the human genome, independent assortment is thought to call into question the assumption that races breed true to type.

The central problem with this argument rests in the assumption that the genes for racial traits must be genetically linked in order for races to breed true to type. Races will breed true to type even when racial traits are independently assorted, provided that there is little outbreeding among racial groups. Inbreeding reduces heterozygosity in the gene pool, thus making it more likely that traits will be passed on faithfully and as a package in reproduction. Cultural factors and geographic separation are two factors that might contribute to reproductive isolation.

Other critics of racial typology argue that races cannot be types because the morphological traits typically used to individuate races (e.g., skin color, hair type, eye shape, etc.) are often clinal and discordant [Banton and Harwood, 1975; Appiah, 1996; Zack, 1997].[9] Not only do many so-called racial traits vary gradually across a geographic region; they also vary independently. For example, a classification scheme based solely on skin color might fail to agree reasonably with one based solely on hair type — and both might fail to agree reasonably with one based solely on bone structure, and so forth. Critics of racial typology sometimes add that a similar problem can be found at the genetic level. Support for the latter claim comes from Lewontin [1972] who argues that there is more genetic variation within than among major racial groups (blacks, whites, and Asians). The conclusion that is frequently drawn is that racial typology is false because these data cast doubt upon the typological assumption that the members of a race have a cluster of properties in common that they do not share with members of other races.

While this argument against racial typology is an improvement over the previous one, it is still not fatal. As Sober has argued [1980], variation within and continuity among taxa need not pose a problem for typological definitions in systematic biology. By appeal to what Sober calls the 'natural state model', pre-Darwinian essentialists had a way of explaining nature's diversity that is consistent with essentialism. According to this model, the members of a kind have a common natural state, determined by the kind-specific essence, but interfering forces often prevent individuals from realizing that state. In other words, essential properties can be thought of as dispositional properties that, when there are no interfering forces at work, have the propensity to manifest the traits typical of the members of a kind. By appeal to this model, then, a racial typologist can recognize significant variation within, and continuity among, races provided that there are discrete natural tendencies underlying this variation.

Not only does this reply reveal a problem with the previous argument, it also points to a better argument against racial typology. It is not variation *per se* that

[9] This argument combines some elements of the previous two arguments against racial typology but also differs from them in that it makes no assumption about the cause of clinal and discordant variation.

poses a problem for typological definitions in systematic biology; the problem is that, with the development of population biology, evolutionary theory no longer uses the natural state model to explain variation within and continuity among taxa [Sober, 1980]. To see why this is so, let us think about the model from the perspective of contemporary genetics. It is the idea that all of the members of a taxon (racial, specific, generic, etc.) share a common genetic essence that, in conjunction with its natural environment, results in the natural phenotype for that taxon. All other phenotypes are regarded as the result of interfering forces and all other environments are regarded as perturbations of the system. The problem is that the ideas of type and of deviation from type do not feature in contemporary evolutionary theory. Contemporary evolutionary theory does not distinguish between the natural phenotype for a given genotype and those that are the result of interference — nor does it identify a natural environment for a given genotype. Moreover, no particular genotype is privileged and viewed as essential for a taxon. Instead all genotypes and environments are treated as being on a par, and the phenotype of a given genotype is always described relative to an environment.

A further problem with typological definitions in systematic biology has been articulated by Michael Ghiselin [1974] and David Hull [1976; 1978]. Their arguments were originally advanced against the typological species concept, but can be generalized to show why typological definitions are problematic at any level in the taxonomic hierarchy. According to Ghiselin and Hull, natural kinds defined according to essentialist criteria are supposed to be spatiotemporally unrestricted classes. They are defined by appeal to purely qualitative properties possessed by all and only the members of a kind. As such, kind membership does not depend on spatiotemporal or causal relations among the members of a kind. Systematic categories, on the other hand, are best understood as spatiotemporally localized historical entities. Most systematic biologists, for example, define taxa (at least in part) in terms of the genealogical relations among organisms. Since genealogical relations are not purely qualitative, Ghiselin and Hull maintain that essentialism is the wrong view to take about biological classification.

It is important to stress that although essentialism was historically influential, not all conceptions of race (common sense or scientific) are essentialist in nature. With the development of contemporary evolutionary theory and the rise of population genetics, the typological race concept was replaced by a population concept of race. It is to this concept that I now turn.

3 THE GEOGRAPHICAL RACE CONCEPT

After the downfall of the typological race concept, biologists began defining 'race' geographically. According to the *geographical race concept*, a race is a geographically localized subdivision of a species that differs phenotypically and genetically from other conspecific populations [Mayr, 1942]. Several features of this definition are worth noting. First, a single geographical race typically comprises several

breeding populations, all of which are very similar. Yet, because there can be slight differences among the populations that make up a race, race membership must be defined with reference to many characters. There must also be statistically significant differences in the mean values of the traits used to define race membership. Second, as a general rule, geographical races are *allopatric populations*; they are populations that are separated from one another by a geographic barrier such as a mountain range, desert, or large body of water. When such populations remain reasonably isolated over time, geographic differentiation is likely to occur. Sometimes, however, geographical races are *sympatric populations* — viz., populations that share the same geographic range but are, in large part, reproductively isolated from other conspecific populations for some other reason. An example of sympatry can be found in some human populations, such as the Amish, that are largely reproductively isolated due to certain cultural and/or religious practices.

The geographical race concept has been used for defining both human and nonhuman race. In its application to humans, it was often assumed that the members of a race are similar to one another, and differ from the members of other races, with respect to certain observable characteristics (skin color, hair type, bone structure, etc). It was also presumed that these traits are heritable and, thus, that there are average differences among the races at the genetic level as well [Dobzhansky, 1953; 1955; Garn, 1961; Coon, 1963]. In the early days of molecular genetics, researchers focused largely on differences in the distribution of blood types (e.g., ABO, Rh, MNS, and Rhesus, etc.) as a measure of underlying genetic differences. Though no group was exclusively of one blood type, biologists found average differences in the distribution of blood groups among human populations from diverse geographic regions. Because such populations are also observably different, a number of biologists and anthropologists took such data as providing support for the existence of human geographical races.

Before discussing some of the problems with the geographical concept, it is worth mentioning that it avoids many of the problems faced by the typological concept. One reason is that, unlike the typological concept, the geographical concept is a population concept. Geographical races are defined in terms of resemblance within and variation among populations, rather than in terms of properties possessed by the individuals that make up a race. A second reason is that, according to the typological concept, kinds are defined by properties possessed by all and only the members of a kind. Though the geographical concept also defines races in terms of shared properties, such properties need not be universal nor unique. As noted above, there must simply be statistically significant differences among the traits used to individuate geographical races. Third, the typological concept requires that the defining properties are intrinsic and explanatory. The geographical concept, on the other hand, simply defines races in terms of clusters of properties without assuming that such properties are due to a common intrinsic causal property (or set of such properties). Indeed, there is often an implicit assumption that external causal factors explain the clustering.

The geographical race concept was eventually abandoned for reasons originally

articulated by E. O. Wilson and W. L. Brown [1953]. Their arguments focused primarily on the use of this concept for defining nonhuman races, but (as we will see in the next section) critics of the biological significance of human race later applied these arguments to the human context. According to Wilson and Brown, the primary problem with the geographical race concept is that the it assumes that geographical variation is typically concordant when, in fact, such variation is often discordant. Racial traits vary concordantly when each of the traits used to define a race yield the same or similar classification schemes. Concordance among a large number of traits is important when individuating geographical races because it indicates reliable covariation in the properties used to define a race. This, in turn, is taken as indirect evidence that the classification scheme is biologically significant.

According to Wilson and Brown, discordant variation has resulted in two problematic trends in systematic biology. As a general rule, discordance increases with an increase in the number of traits used to individuate a race. Consequently, some systematists began naming geographical races on the basis of a small number of traits — and, thus, were arbitrarily ignoring nonconforming characters. This is problematic because, without reliable covariation among a large number of traits, there is no reason to suppose that the resultant designations are biologically significant. Other taxonomists tried to circumvent the problems caused by discordant variation by naming races that are restricted to very small geographic areas. One problem, here, is that such designations result in an overly detailed classification scheme that is of limited use to the systematic biologist. A second difficulty is that, because the most concordant infraspecific units are frequently local breeding populations, the term 'race' becomes synonymous with 'breeding population' and, thus, loses its significance as a concept.

In addition to the problems discussed by Wilson and Brown, there is an underlying theoretical problem with the geographical race concept. The geographical concept is a phenetic concept [Andreasen, 2000]. Pheneticists define taxa in terms of the overall similarity of their members. Breeding populations are grouped into races based on phenotypic and genetic similarities; races are grouped into species by the same method, and so on up the taxonomic hierarchy. A central problem with pheneticism is that for any population or set of populations, there are frequently several competing patterns of similarity that could be used to define a taxon [Hull, 1970; Ridley, 1986; 1993; Sober, 1993]. Pheneticism in general, and the geographical race concept in particular, offers no nonarbitrary way of picking one similarity grouping as the correct one.

It is now widely agreed that both the typological and geographical race concepts are unacceptable. One might be tempted to conclude from their respective failures that human races are biologically unreal. Yet, such a conclusion would be too hasty. The arguments discussed above should be thought of as *local* arguments against the biological reality of both human and nonhuman race; each aims to show that some particular biological race concept is unacceptable. The arguments that I will discuss in the next section are, in a certain sense, more narrow. Each argues

for the nonexistence of *human* biological races. In a different sense, however, they are more *global*. Rather than focusing on a particular race concept, each aims to explain why *any* proposed biological definition of human race is doomed to failure.

4 GLOBAL ARGUMENTS AGAINST HUMAN BIOLOGICAL RACES

In this section I will present three global arguments against the biological reality of human race.[10] I will call the first 'the no subspecies argument', the second 'the independent variation argument', and the third 'Lewontin's genetic argument'. These arguments have been extremely influential and are widely assumed to prove, unambiguously, that human races are biologically unreal. Yet, as we will soon see, these arguments are not as conclusive as people often suppose.

Let us begin with what I have called 'the no subspecies argument'. Defenders of this argument aim to show that human races are biologically unreal by establishing that there is no respectable race concept in systematic biology [Livingstone, 1964; Gould, 1977; Futuyma, 1986]. This argument starts with the assumption that the term 'race' is synonymous with the taxonomic term 'subspecies'. Human races, for example, are subspecies of *Homo sapiens*. Next, it is argued that the subspecies concept has been discredited in systematic biology. After the downfall of the typological and geographical concepts, many systematists simply gave up the practice of dividing species into subspecies. One reason is that some began to worry that the subspecies concept is misleading. On their view, it suggests discrete units of variation when, in reality, infraspecific variation is often clinal and discordant. A second reason is that there are other ways of studying variation below the species level — and, thus, some have argued that the subspecies concept is superfluous. Proponents of this argument conclude that if there is no legitimate subspecies concept in systematic biology, then there is no legitimate basis for the everyday practice of dividing humans into biological races.

The central problem with this argument is that the subspecies concept has not been fully abandoned in systematic biology [Kitcher, 1999; Pigliucci and Kaplan, 2003; Andreasen, 2004; Gannett, 2004].[11] After the downfall of the typological and geographical concepts, some systematists began defining subspecies *phylogenetically* — as distinct evolutionary lineages within a species [Templeton, 1999; Shaffer and McKnight, 1996; Legge *et al.*, 1996; Miththapala, 1996].[12] As we

[10] The distinction between global and local arguments against the biological reality of race is not frequently made in the race literature. Indeed, sometimes an author draws a global conclusion from a local argument. For these reasons, I have reconstructed some of the arguments that I will present in this section.

[11] An important difference between Gannett, on the one hand, and Kitcher, Andreasen, and Pigliucci and Kaplan, on the other, is that Gannett rejects the biological reality of human race whereas Kitcher, Andreasen, and Pigliucci and Kaplan suggest that human biological races might exist (or might have once existed).

[12] Although Templeton allows for the possibility that subspecies might be defined phylogenetically, he denies that *human* races can be defined in this way. Part of the reason is that he believes that there is, and always has been, too much outbreeding among human populations for

will seen in the next section, a number of different phylogenetic conceptions have been offered. What is important for our purpose, however, is that they all differ from the typological and geographical concepts by requiring that races be distinct evolutionary lineages. Others, especially plant systematists, turned their attention to an *ecological* subspecies concept [King and Stansfield, 1990; Stone *et al.*, 2001; Vincente *et al.*, 2001].[13] On this view, races are infraspecific groups that have become genotypically, and often phenotypically, differentiated as a result of differential selection pressures from different local environments. This concept is similar to the geographical race concept in that it defines races, at least in part, as phenotypically and/or genotypically distinct populations. Yet, unlike the geographical concept, the ecological concept allows that races can be defined on the basis of very few characteristics [Pigliucci and Kaplan, 2003]. A further difference is that the ecological concept does not require geographic localization. It simply requires phenotypic and/or genetic differentiation that is due to a common selective regime. Not only is it possible for similar ecotypic characteristics to evolve independently in distinct geographic locations; it is also possible for two or more distinct ecological races to coexist in the same geographic location. I will have more to say about each concept below. What is important for now is that, in order for the 'no subspecies argument' to be successful, one would need to argue that the phylogenetic and ecological race concepts have been — or ought to be — discredited.

Defenders of the second and third global arguments, by and large, aim to show that there can be no adequate biological definition of human race by arguing that biology lends no support to common sense conceptions of race. An implicit assumption at work, here, is that any biological race concept, were it to deviate too far from common sense, is not really a concept of 'race'. One difficulty that must be addressed, however, is that the term 'race' often gets used in a number of different, sometimes conflicting, ways in everyday discourse [Smedley, 1993; Omi and Winant, 1994; Wright, 1994]. Critics of the biological race concept are aware of this problem but argue that, in spite of the differences, there are certain core elements common to all or most common sense conceptions of human race [Root, 2001; 2003; Zack, 2002; Keita and Kittles, 1997].[14,15] Two such elements

phylogenetic races to exist.

[13]According to Pigliucci and Kaplan [2003], some systematists combine these concepts and, thus, define subspecies ecologically and phylogenetically.

[14]Hardimon [2003] also makes this assumption, but does not use it to argue against the biological reality of human race. His aim is simply to characterize what he takes to be the everyday notion of race. Likewise Hirschfeld [1996] as well as Machery and Faucher [forthcoming] make this assumption in the context of their discussion of the possibility that such elements are due to some universal aspect of human cognition. Like Hardimon, however, they do not use this as part of an argument against the biological reality of human race.

[15]It is unclear whether, and to what extent, defenders of this type of argument take these elements to be universal across cultures and historical periods. On the one hand, race constructionists often maintain that there is a diversity of race concepts that covary with cultural and historical differences. On the other hand, defenders of this type of argument rarely relativize their claims to a specific time or place.

will be important for understanding the arguments in this section. First it is often said that common sense demands that races be defined, at least in part, in terms of clusters of directly observable properties — and, in particular, that there be reasonable continuity within, and significant difference among, the races with respect to such properties [Root, 2001; 2003; Zack, 2002; Keita and Kittles, 1997; Keita *et al.*, 2004]. Although there is some disagreement over which properties are supposed to be the defining ones, skin color is frequently taken to be central. Other traits such as hair type, eye shape, and bone structure are often included as well. Second, it is often said that common sense requires that such features be heritable and, thus, that the members of a race be reasonably genetically similar to one another and significantly genetically different from the members of other races [Lewontin, 1972; Root, 2001; 2003; Zack, 2002; Keita and Kittles, 1997; Keita *et al.*, 2004].

With these ideas in the background, I will now present 'the independent variation argument'. This argument takes its lead from Wilson's and Brown's main objection to the geographical race concept. Recall that, according to Wilson and Brown, the geographical concept is problematic because it assumes that phenotypic variation is often discrete when, in fact, it is typically discordant. Analogously, critics of biological conceptions of human race sometimes argue that it is not possible to provide a biologically respectable definition of human race because biology lends no support to the common sense idea that races ought to be defined in terms of clusters of covarying observable characteristics. On the contrary, it is argued, most of the traits typically used to individuate races vary independently. The idea at work here is that a classification scheme based on skin color, for example, might cross-classify one based on hair type — and both might disagree with one based on eye shape (etc.). Because this problem is said to be compounded as more traits are added to the classification scheme, defenders of this type of argument conclude that human races are biologically unreal [Diamond, 1994; American Anthropological Association, 1998; Root, 2001; 2003; Zack, 2002].

Some philosophers have argued, however, that the above argument places too much emphasis on the idea that races ought to be defined in terms of clusters of overt physical features [Andreasen, 1998; 2000; Kitcher, 1999]. Biologists and philosophers have known for some time that overall similarity, although sometimes useful, can be misleading as a method for uncovering a stable classification scheme [Goodman, 1978; Ridley, 1986; 1993; Hull, 1970; 1998]. In addition, the assumption that races must form discrete phenotypic clusters is an implicit demand for essentialism [Pigliucci and Kaplan, 2003; Sesardic, 2003]. For these reasons, it is not clear that we ought to hold biology to this particular common sense belief. Furthermore, discordance among overt physical features may not, in fact, be a violation of common sense. Independent variation is a matter of degree — and presumably common sense allows for some discordance among overt physical features. Of course, a defender of the independent variation might respond that there is *too much* discordance, and thus too much deviation from common sense,

for human races to be biologically real. However, it is hard to know how much discordance is too much, and little effort has been taken to address this question.

The final global argument that I will discuss was originally advanced by Richard Lewontin [1972], but has been repeated many times since its original formulation [American Anthropological Association, 1998; Root, 2001; 2003; Zack, 2002; Keita *et al.*, 2004]. This argument parallels the independent variation argument, but does so at the genetic level. It is the idea that biology lends no support to the common sense assumption that for races to be biologically real there must be significant genetic similarity within, and significant genetic difference among, the major racial groups. Defenders of this argument maintain that genetic studies reveal that there is more genetic variation within than among major racial groups. Lewontin [1972, 396], for example, states: "The mean proportion of the total species diversity that is contained within populations is 85.4% ... Less than 15% of all human genetic diversity is accounted for by differences between human groups! Moreover, the differences between populations within a race accounts for an additional 8.3%, so that only 6.3% is accounted for by racial classification."[16] Other geneticists have reported similar findings, and though the numbers are somewhat different, the overall pattern is said to be the same.

Anthony Edwards [2003] has provided an interesting and often overlooked challenge to Lewontin's argument. Edwards' objection can be broken into two parts. The first aims to establish that Lewontin's analysis of his data, although fine for some purposes, is not the right sort of analysis for addressing questions about classification. Lewontin examined the relative frequency of 17 polymorphic proteins in 7 populations (Caucasian, African, Mongoloid, S. Asian Aborigines, Amerins, Oceanians, Australian Aborigines) and bases his argument on a locus-by-locus analysis of these data. He, thus, overlooks (or ignores) the fact that gene correlations provide information in addition to gene frequencies. This is problematic, according to Edwards, because it is often the correlation structure of the data that enables the discovery of a stable classification scheme. Edwards concludes that Lewontin's argument is circular. By relying on gene frequencies alone, he ignores the structural aspects of his data and then concludes that these data possess no such structure. The second part of Edwards' argument aims to establish that once the right type of analysis is applied, there is likely to be more genetic differentiation among human populations than is commonly supposed. This goes beyond the purely methodological point discussed above.

So far we have seen some problems specific to each global argument. In addition to these, there are at least two problems shared by the second and third global arguments. Both of these problems turn on the shared background assumption that it is possible to prove the nonexistence of human biological races by showing that biology lends little support to certain core features of everyday conceptions of race. The first problem is that the question of what everyday folk mean by

[16]For the purpose of this study, Lewontin identified the seven racial groups listed above. However, elsewhere [1984] he argues that the results are robust and hold up no matter how races are defined.

race is an empirical question. Empirical data can come from a number of different sources — including historical uses of the term, references to 'race' in governmental documents, psychological studies on how people think about race, sociological surveys, and so forth.[17] Yet defenders of the second and third global arguments often simply stipulate the features that they take to be central to common sense, without providing empirical support for their claims.[18] Without such support, however, it is hard to know why we should assume that the core features specified above are so central that failure to meet them constitutes a global argument against the biological reality of human race. Indeed, because the term 'race' has taken on a number of different meanings throughout history, we should not overlook the possibility that there is no single set of core features possessed by all or most common sense conceptions of race.[19] It is not implausible to suggest, for example, that a number of different ideas have been dominant in different cultures and/or different historical periods without there being any single feature or set of features shared by all or most everyday conceptions of race.

A second difficulty is that it is not clear that we should practice complete semantic deference to "the" everyday notion of race. Let us suppose for the sake of argument that the two "core" elements specified above form an important part of the everyday race concept. Let us also suppose that these assumptions are biologically unjustified. What conclusion should one draw, if both of these claims are true? Defenders of Lewontin's genetic argument and the independent variation argument suggest that we ought to conclude that races are biologically unreal. But why should we accept this conclusion? Suppose that one were to find a biologically significant conception of race that deviates somewhat from the features specified above. Why not say that there is a biologically respectable definition of 'race' that deviates somewhat from common sense [Andreasen, 1998; 2000; 2005; Pigliucci and Kaplan, 2003]? In response to this type of question, some have argued that such a conception would not count as a concept of 'race' because it deviates too far from everyday uses of the term [Zack, 2002; Glasgow, 2003]. I will briefly respond to this worry in the next section. For now, let me make a different point. If there are core elements common to most everyday notions of race, the two elements specified above need not be the only ones. Common ancestry and/or geographic location, for example, have also been cited as central features of the everyday race concept [Hardimon, 2003; Root, 2001; 2003]. It is possible, therefore, that there is a biologically respectable definition of 'race' that agrees with these features, even if such a definition disagrees with the features specified above.

[17] See, for example, Gould [1981]; Smedley [1993]; Wright [1994; 1995]; Appiah [1996]; Hirschfeld [1996]; Hirschman [2000]; Machery and Faucher [forthcoming].

[18] Examples include Root [2001; 2003], Zack [2002], and Hardimon [2003], though Hardimon does not take a stance on the biological reality of race.

[19] See Hirschfeld [1996] and Machery and Faucher [forthcoming] for the opposing idea that there is a universal cognitive mechanism that explains certain patterns in everyday reasoning about race.

5 ECOLOGICAL AND PHYLOGENETIC CONCEPTIONS OF RACE

The three global arguments discussed in the previous section have been highly influential and have played an important role in establishing what many take to be the received view about race — viz., that human races are biologically unreal. As we have just seen, however, not everyone takes them to be decisive. Indeed, doubt about their success has led some philosophers and biologists to revisit the question of the biological reality of human race and to ultimately argue for what they take to be improved biological conceptions of race.

As noted earlier, two types of accounts have been offered — one ecological and another phylogenetic. Let us recall that ecological conceptions of race typically define races as subspecific groups composed of individuals who are phenotypically and genetically similar to one another due to a common selective regime. Phylogenetic conceptions, on the other hand, define races in part as lineages of reasonably reproductively isolated breeding populations. I will discuss these views in more detail momentarily. For now, I would like to note that such accounts are often taken to be improvements over their predecessors, in part, because they are relatively minimalist in comparison with previous biological conceptions. For example, each calls into question the racialist assumption that knowledge of a person's race allows one to make inferences about the psychological and behavioral traits that she is likely to possess.[20] Second, each abandons the idea that races must form discrete genotypic and/or phenotypic clusters for races to be biologically real. Consequently, each allows that human biological races might exist despite the empirical findings typically cited against the biological reality of human race.[21] In what follows, I will discuss a number of difficulties that the ecological race concept must address. I will then defend phylogenetic conceptions of human race against three important objections.[22]

The ecological race concept, as well as its application to humans, has recently been endorsed Massimo Pigliucci and Jonathan Kaplan [2003]. Like the geographical race concept, the ecological race concept defines a race, in part, as a group of individuals who have a number of phenotypic and genetic similarities in common. Yet, as noted earlier, there are also a number of important differences. The ecological race concept requires that the similarities and differences used to individuate races be *adaptive* similarities and differences, whereas the geographical race concept has no such requirement. Second, the geographical concept requires that race membership be defined with reference to many characteristics. Pigliucci and Kaplan, however, allow that ecological races can be named on the basis

[20] Pigliucci and Kaplan [2003] are more willing than either Kitcher [1999] or Andreasen [1998; 2004] to acknowledge the possibility that knowledge of a person's race might allow one to make inferences about her psychological and behavioral traits.

[21] See Pigliucci and Kaplan [2003] and Andreasen [2004] for more developed discussions of this point.

[22] For more detailed defenses of phylogenetic conceptions of race, see Andreasen [1998; 2000; 2004; 2005], Kitcher [1998], and Risch *et al.* [2002].

of very few characteristics.[23] Third, geographical races are typically allopatric populations, while ecotypic races will often be sympatric. Not only can distinct ecological races coexist in the same geographic region, similar ecotypes can and do evolve independently in diverse geographic locations.

To further enhance our understanding of the ecological race concept, let us briefly examine how it differs from phylogenetic conceptions of race. The main difference between phylogenetic conceptions of race and the ecological race concept is that the former require that races be lineages of human breeding populations, while the latter has no such a requirement. Indeed, Pigliucci and Kaplan maintain that different ecologically important traits may sometimes vary independently. Thus, on their view, it is possible for a single breeding population to belong to different ecological races. In addition, Pigliucci and Kaplan argue that ecotypic races can exist despite significant gene flow — so long as the forces of selection are sufficient to maintain the genetic differences that cause the different adaptive phenotypes. This difference, according to Pigliucci and Kaplan, is one the main advantages that the ecological race concept has over phylogenetic conceptions of race. Against the phylogenetic conceptions of race, it is sometimes argued that races cannot be lineages because there is, and always has been, too much gene flow among human populations for human phylogenetic races to exist. I will discuss this objection, as well as some responses to it, in more detail below. For now it is important to note that, according to Pigliucci and Kaplan, the ecological race concept avoids this difficulty because ecological races need not be phylogenetic units.

Though Pigliucci and Kaplan present little data in support of their view (a point that I will return to momentarily), they do provide a couple of suggestive examples that illustrate its application to humans. They maintain, for example, that skin color is an ecologically significant trait and, thus, implicitly suggest that there are ecological races associated with skin color differences (and possibly other adaptive phenotypes as well).[24] They also discuss the possibility that there might be ecotypic races associated with regional differences in physiology and, hence, athletic ability. For example, West African populations produce more world-class sprinters, on average, than other populations from around the world. Likewise, Kenyan populations produce more successful marathon runners, on average, than other populations from around the world. Pigliucci and Kaplan [2003, 1167] hold that if such regional differences in the production of top athletes reflect regional differences in physiology — and if such differences are adaptations to different local environments — then there would "indeed be races associated with athletic ability".

The above examples are both intuitive and appealing. Nonetheless, there are a number of problems that need to be addressed before one could reasonably conclude that human races are ecotypes. The first problem is one that I alluded to

[23] It is unclear from what they say whether this is a widely held view about ecotypes.

[24] It is unclear, given what Pigliucci and Kaplan say, whether skin color alone is sufficient to designate ecotypic races.

earlier — namely, there is little empirical support for the application of this concept to humans. The idea that human races are ecotypes is a conceptual possibility and making note of this fact is, in and of itself, a valuable contribution to the race debate. Nonetheless Pigliucci and Kaplan make a stronger claim — i.e., that human ecotypic races exist — but provide no data and few examples to support their view.

Second, the suggestion that human races and human ecotypes are one and the same requires argument. On the surface, at least, these concepts are importantly different. Part of the reason is that the two terms — 'race' and 'ecotype' — are typically used somewhat differently in common sense. As Pigliucci and Kaplan themselves argue, human ecotypes need not (and often will not) correspond with folk racial categories. Folk racial categories are often said to be based, at least in part, on ancestral geographic origins. For example, a person who is classified as 'black' is frequently assumed to have African ancestral origins. Yet, according to Pigliucci and Kaplan, a classification system based (in part) on geographic origins need not, and often will not, correspond with one based on adaptive phenotypes alone. On their view, not only is it possible for the same or similar ecotypic traits to evolve independently in different geographic locations — thus allowing for sameness of race across diverse geographic regions — it is also possible for different ecological races to evolve in the same geographic location. A second reason why there might be an extensional mismatch between common sense conceptions of race and the ecological race concept stems from the widely held assumption that common sense conceptions of race require concordant variation in racial phenotypes. (It is this assumption that is at work in the argument that I have called 'the independent variation argument'.) Pigliucci and Kaplan, however, argue that independent variation is not a problem for their view. In fact they argue that, due to independent variation in adaptive phenotypes, a single population can belong to more than one ecotypic race.

Perhaps in partial response to this type of worry, Pigliucci and Kaplan maintain that their account is not "completely orthogonal to folk conceptions of race" [2003, 1166]. However, they fail to explain why they think that this is so — and , because they provide no data and few examples of human ecotypic races, it is hard to know exactly what they have in mind. Furthermore, they are not consistent on this point. Sometimes they maintain that human ecotypic races are "not completely original" to folk conceptions of race — but at other times they maintain ecotypic races have "little or nothing in common with folk races" so much so that they suggest that we avoid the term 'race' altogether [2003, 1161-2].

Of course, one need not hold science hostage to common sense. Regardless of how the terms are used in everyday discourse, one might argue that the terms 'race' and 'ecotype' *should* be treated as synonymous, if they are used interchangeably in most scientific contexts. The problem, however, is that these terms are not always used interchangeably in science — especially when talking about human race. While it is true that these terms are sometimes used interchangeably in some branches of systematics — for example, some systematists identify ecological

subspecies in plants and insects — this is not how scientists in general typically think of human race. By and large biologists and anthropologists — whether they defend or deny the biological reality of human race — have either assumed one of three types of definitions. They have either assumed that human races ought to be defined by appeal to race-specific essences, that they ought to be defined as groups of geographically localized individuals with a number of phenotypic and genetic properties in common, or that they ought to be defined, at least in part, as lineages. As we have already seen, Pigliucci and Kaplan maintain that their account is importantly different from each of these conceptions of race. These factors, taken in conjunction with the worries expressed above about the relation between 'race' and 'ecotype' in everyday discourse, leaves them open to the demand for an explanation as to why one ought to suppose that human races are ecotypes.

A final difficulty with Pigliucci's and Kaplan's ecological race concept is that it lacks development. Although there are certainly adaptive differences among distinct human populations, it is not clear when and under what conditions such differences constitute different ecotypic races. Does each adaptive difference constitute its own racial division, or do only certain adaptive differences — or certain sets of adaptive differences — count? Since ancestry and geographic location are two ideas commonly associated with race, one might be tempted to rely on these factors in an effort to explain which adaptations are the racial ones. Yet, because Pigliucci and Kaplan argue that there is and always has been too much gene flow for human races to be distinct lineages, they cannot use phylogeny to answer this question. Likewise, due to some of the ways in which their account differs from the geographical race concept, they cannot rely on geographic location for an answer. To further complicate matters, let us recall that Pigliucci and Kaplan maintain that human ecotypic races can, and often will, deviate significantly from folk racial categories. This statement, in conjunction with their claim that that ecotypic races can be named on the basis of very few characteristics, leaves the reader wondering how many ecological races there are and which groups constitute ecological races. In fact, when one recalls that a single population can belong to more than one ecotypic race, it appears that the number of races could be very large indeed. At the extreme, it is possible that each individual or population could belong to its own ecotypic race — thus trivializing the concept of race.

I have just discussed several issues that must be addressed before one can reasonably conclude that human ecotypic races exist. Let us now turn to some phylogenetic conceptions of race. Recall that although there are a number of different phylogenetic conceptions of race, all of them assume that races ought to be defined, at least in part, as reasonably reproductively isolated breeding populations. Robin Andreasen [1998; 2000; 2004; 2005] and Philip Kitcher [1999] have independently defended somewhat different phylogenetic conceptions of race. After briefly discussing each, I will defend both against some of the central objections advanced against the idea that races are phylogenetic units.

Andreasen has advanced and defended a view that she calls 'the cladistic race concept'. Cladistics is a branch of systematic biology that, with reference to a well-

confirmed branching structure, defines taxa solely in terms of common ancestry. Cladistic classification was originally developed by Willi Hennig [1966] for defining higher taxa. On his view, higher taxa are monophyletic groups (groups composed of an ancestor and all of its descendants) of well-defined species.[25] Andreasen argues that, on the assumption that it is possible to represent human evolution as a branching process, cladistic classification can be adapted for defining race. Cladistic races would, thus, be monophyletic groups of reasonably reproductively isolated breeding populations.[26]

Next, Andreasen argues that some support for the existence of human cladistic races can be found in current work in human evolution. In particular, a number of different research groups have proposed branching diagrams that aim to represent human evolutionary history — i.e., patterns of migration and subsequent geographic isolation among human breeding populations.[27] Andreasen maintains that, assuming that such diagrams are (or can be) empirically well-supported, they suggest that the condition of reasonable reproductive isolation has been met in the distant past. This, in turn, suggests on her view that human cladistic races once existed. Yet, because Andreasen also speculates that it is unlikely that the condition of reasonable reproductive isolation is being met today, she concludes that human races are most likely on their way out.

Like Andreasen, Kitcher [1999] argues that races ought to be defined phylogenetically. On his view, races are founder populations and inbred lineages that have become differentiated (phenotypically or genetically) in the absence of reproductive contact. The idea at work, here, is that racial divisions begin with the division of a species into founder populations, possibly due to migration or some other form of geographic separation. Over time, and with limited interbreeding among geographically separated populations, phenotypic and genetic differentiation is likely to occur. Kitcher argues that such differentiation can be used for individuating races — provided that when previously separated populations are brought back together, there is enough reproductive isolation to sustain the distinctive phenotypic or genetic properties that mark the races. One way for this to occur in humans is if previously separated populations develop certain cultural barriers to outbreeding.

In defense of the application of his view to humans, Kitcher argues that data on the geographical history of human populations suggest that, in the distant past, there was most likely sufficient geographical separation and reasonable re-

[25]It is important not to confuse cladistic classification with cladistic methods of phylogenetic inference [Sober, 1993]. Doing so will result in a number of faulty arguments against the cladistic race concept (see, for example, [Zack, 2002]).

[26]There is, of course, some ambiguity in the idea of reasonable reproductive isolation. Part of the reason is that, as members of the same species, all humans can interbreed. Nonetheless there can still be significant reproductive isolation among human groups — either due to sociocultural barriers or due to geographic isolation. See Kitcher [1999] for further development of the concept of reproductive isolation and its relation to racial classification.

[27]See, for example, Nei and Roychoudhury [1972; 1974; 1982; 1993]; Vigilant *et al.*, [1991]; Wilson and Cann [1992]; Cavalli-Sforza *et al.*, [1994]; Risch [2002], Bamshad and Olson [2003].

productive isolation among human breeding populations. In addition to these data, Kitcher also relies on data on the rates of interracial relationships and reproduction in the United States today. While admitting that such data are limited, Kitcher argues that those data that do exist suggest that there is still reasonable reproductive isolation among human breeding populations today. Thus, he tentatively concludes that some biologically significant races exist in the United States today.[28]

Let us compare and contrast the cladistic race concept with Kitcher's conception of race. Both accounts define races historically — i.e., as lineages of reasonably reproductively isolated breeding populations. Yet, on the cladistic account, genealogy alone is sufficient for defining race, whereas Kitcher's model treats genealogy as a necessary, but not sufficient, condition. In addition to genealogical differentiation, his model also requires that (A) there be some phenotypic or genetic distinctness among distinct races, and that (B) the residual mixed race populations be relatively small. Second, because the cladistic account requires that races be monophyletic groups, a population must be reproductively isolated over a significant portion of evolutionary history before it can be designated a cladistic race. Kitcher's account, on the other hand, does not require monophyly. On his view, races may sometimes be historical (diachronic) lineages, but sometimes we may want to recognize nondimensional races — viz., "groups at a particular place at a particular time that are not exchanging genes at substantial rates" [1999, 243]. Third, Andreasen and Kitcher provide different kinds of support for their views. For the most part, Kitcher uses data on the rates of interbreeding among major racial groups in the U.S. today. Andreasen, on the other hand, uses current work in human evolution. This leads to a fourth, and perhaps the most important, difference between the two views. Kitcher is more optimistic than Andreasen about the existence of human races today.

As noted earlier, one of the primary objections to phylogenetic conceptions of human race is that there is, and always has been, too much gene flow among human breeding populations for distinct lineages to have evolved [Wolpoff *et al.*, 1988; Wolpoff, 1989; Thorne and Wolpoff, 1992; Templeton, 1999]. For example, one way of understanding the genetic data cited by Lewontin and others (in what I have called 'Lewontin's genetic argument') is to see it as support for the claim that human populations were never been reasonably reproductively isolated over a significant portion of evolutionary history. Pigliucci and Kaplan [2003, 1164-1165] summarize this position well:

> The current distribution of genetic variation within *H. sapiens* implies that at no time in the past were any of the (currently extant portions of the) population evolving independently. While the *Homo* genus very likely generated incipient species during its history, none of these currently survive. The evolution of contemporary *Homo sapiens* was likely *not* marked by populations that at one time had independent evolu-

[28] The racial groupings that Kitcher has in mind are blacks, whites, and Asians.

tionary trajectories but exist today as part of the larger population.

Though this objection applies to Kitcher's account as well as to the cladistic account, one might argue that it is particularly problematic for the cladistic race concept. One reason why cladistic classification is typically applied at or above the species level is because monophyly requires evolutionary branching without reticulation. Since significant gene flow prevents branching, it would thus impede the formation of cladistic races [Zack, 2002; Gannett, 2004].

This is probably the strongest objection to phylogenetic conceptions of human race — yet, it is not fatal. Part of the reason is that, as noted above, there is some disagreement over the significance of Lewontin's data. Edwards [2003], for example, questions Lewontin's analysis of his data on the grounds that it fails to take gene correlations into consideration. In addition, he adds that once gene correlations are taken into consideration, it is possible to derive a convincing branching diagram of human evolutionary history. Like Edwards, many other human geneticists accept Lewontin's analysis of his data, but nonetheless maintain that the allele frequencies differences that exist are highly structured and are useful for identifying distinct lineages [Prichard *et al.*, 2000; Risch, 2002; Rosenberg, 2002; Bamshad *et al.*, 2003]. Taking a different (but consistent) approach, Andreasen [1998] has suggested another reading of Lewontin's data. On her view, the data are consistent with the possibility that, while there may be significant outbreeding today, there was once enough reproductive isolation for distinct lineages to have evolved. Indeed, this is the main reason why she suggest the possibility that human races once existed, but are now anastamosing.

That being said, I would like to add that Andreasen and Kitcher each maintain that their thesis is best understood as a conditional claim: If there is, or has been, reasonable reproductive isolation among human breeding populations, then it is possible to provide a biologically significant definition of human race. Both are aware of competing models of human evolution — such as the trellis model or the multiregional evolution hypothesis — which assume that there is, and always has been, too much gene flow among human populations for phylogenetic differentiation to have occurred [Wolpoff *et al.*, 1988; Wolpoff, 1989; Thorne and Wolpoff, 1992; Templeton, 1999]. Furthermore, both acknowledge that if such a model were to be empirically vindicated, then human races don't exist and never have existed.

A second objection to phylogenetic conceptions of race has been advanced by Lisa Gannett [2004]. Gannett's objection has three-parts.[29] Her main concern is that the branching diagrams (constructed by defenders of candelabra models of human evolution) upon which phylogenetic conceptions rest are not empirically well supported. They are not empirically well supported, on her view, because they lack independent confirmation that the breeding populations used in tree

[29]Gannett's objection focuses largely on the cladistic race concept, but since much of what she says applies to other phylogenetic conceptions as well, I will present her objection as being more general.

reconstruction "have not simply been imposed on evolutionary history". Citing Alan Templeton, a defender of the trellis model of human evolution, she suggests that defenders of candelabra models have overstepped their bounds and that they have merely *assumed* that human evolution takes the form of a branching diagram.[30] Templeton states [1999, 639]: "The computer programs used to generate "trees" from genetic distance data will do so regardless of what evolutionary factors generated the distances. It is therefore the obligation of the users of such programs to ensure that the genetic distance data have the properties of treeness before representing their data as a tree". To this Gannett adds that by defining races and breeding populations in terms of relations among groups, rather than among individuals, there is further encouragement for assuming "the property of treeness". Finally, she maintains that Andreasen has ignored the warnings of feminist epistemologists that scientific investigation is inherently value-laden and that theory construction and confirmation cannot be free from bias. She concludes, therefore, that there are good reasons to question the objectivity of the branching diagrams upon which the cladistic race concept rests. Thus, on her view [2004, 324], the cladistic race concept — and by extension other phylogenetic conceptions of race as well — constitutes "an illegitimate biological reification of race".[31]

Let me begin by responding to Gannett's worry that a branching pattern has been imposed on evolutionary history and that there is little support for the assumption of treeness. First, it is important to note that Templeton's objection applies only to phylogenetic trees constructed using genetic distance based methods. Since other methods of inferring population structure exist, his argument is somewhat limited in scope. Using model-based methods, for example, Noah Rosenberg *et al.* [2002, 2382] and others [Prichard *et al.*, 2000; Mountain *et al.*, 1997] have argued that analysis of individual multilocus genotypes permits researchers to infer ancestry without relying on information about sampling locations of individuals. Rosenberg and his colleagues tested the correspondence of predefined populations (defined, for example, on the basis of culture or geography) with groups that have been inferred from individual multilocus genotypes. They used a computer algorithm that clusters individuals with distinctive allele frequencies. Clusters were made blind, meaning that genetic data were fed into the computer without including information about the populations from which these data came. Rosenberg, et al. found a general agreement between predefined populations and those that were inferred from individual multilocus genotypes.

[30] According to this trellis model, there has always been significant gene flow among human evolutionary history. Thus, defenders of this model deny that human evolution takes the form of a branching diagram.

[31] In addition to this three-part objection, Gannett maintains that the cladistic concept ignores racial admixture. This is a misrepresentation of the Andreasen's view — since she holds that distinct lineages once existed, but that human are now anastamosing. Another misrepresentation occurs when Gannett describes the cladistic concept as a static conception of race. Andreasen clearly states that races are dynamic categories. Finally, Gannett refers to the cladistic concept as a "false typology". Although she states that she does not have essentialism in mind, the use of the term 'typology' is politically charged and misleading.

In response to Gannett's other two worries — viz., her conceptual worry that defining races in terms of groups encourages the assumption of treeness as well as her worry that ideological bias is at work — I would simply like to note that an empirical argument is needed in both cases. Take the ideological claim, for example. One would need to demonstrate *empirically* that bias plays a significant role in the confirmation of human phylogenies and that there is too much bias for such theories to be objective. It would also need to be shown that the bias is due to the goal of advancing some political agenda.[32] Indeed, though most philosophers of science agree that science is a social process, there are important differences of opinion among philosophers of science (feminist and non-feminist) about the extent to which ideological bias is filtered out in the process of confirmation.[33]

Related to her charge of bias, Gannett also claims [2004, 331] that Andreasen's account "is problematic in its maintenance of rigid dichotomies between science and society, facts and values, nature and culture, and the biological and the social". Such dichotomies are problematic on her view because they leave "no room to inquire about ways in which biological and social forces interact in the structuring of human groups". I disagree with Gannett on this point Indeed, as Kitcher [1999] has argued, social and cultural factors can produce one mode of reasonable reproductive isolation. As for endorsing rigid dichotomies, the only dichotomies that Andreasen is committed to are ones between certain biological conceptions of race and some social constructionist conceptions and ones between certain sociological questions asked by some social constructions and certain biological questions asked by some human evolutionists. This does not entail, however, that there is no room to inquire about the role of the social.

The final objection to phylogenetic conceptions of race that I would like to discuss allows for the possibility that distinct lineages exist, or once existed, within *Homo sapiens*. Defenders of this objection maintain that such lineages are not races, however, because they deviate too far from common sense conceptions of race [Zack, 2002; Glasgow, 2003]. Joshua Glasgow [2003] has provided the most detailed formulation of this objection. Glasgow focuses primarily on the cladistic concept, because he thinks it suffers the most difficulties, but also argues that Kitcher's view suffers similar difficulties.

Glasgow argues that there is significant extensional disagreement between the cladistic concept and common sense conceptions of race. For example, he asserts that the number of races recognized by the cladistic concept, which he assumes is nine, exceeds the number recognized in common sense. While admitting some disagreement within common sense, he holds that rarely does the number exceed five. Glasgow also reminds the reader that the cladistic concept raises the possibility that 'Asian' may not be a cladistic race — and, thus, that the cladistic concept cross-classifies common sense racial groupings. Next, Glasgow argues that there is significant intensional disagreement between common sense and the cladistic

[32] I am, here, employing an argument originally discussed by Sober [1993] in reference to the charge that sociobiology is ideological.

[33] For more on this issue, see Longino [1990] as well as the *Monist*, volume 77, 1994.

race concept. The main problem, here, is that skin color (and often other overt features as well) is (are) often an inextricable part of common sense conceptions of race. Yet, according to Glasgow, the cladistic concept treats overt morphology as irrelevant to race because it defines races in terms of genealogy alone. After considering and rejecting the possibility of semantic deference to science over the meaning of 'race', Glasgow concludes that cladistic races are not really 'races'.

Andreasen [2005] responds to Glasgow in part by arguing that he tells a selective story about the meaning of 'race'. She adds that once we examine a fuller picture, we see more overlap — both intensional and extensional — between the cladistic concept and common sense than Glasgow allows. She adds that because Glasgow allows that minor revisions in meaning are allowable, this is sufficient to show that the cladistic race concept is legitimately a theory about 'race'. Nonetheless, because she acknowledges that some divergence between the cladistic concept and common sense is possible, she argue that some cross-classification is not a problem for the cladistic view. Part of the reason is that differing folk conceptions of race sometimes cross-classify on another as well.[34]

Ultimately, the disagreement between Glasgow and Andreasen is over how to settle the meaning of purported natural kind terms when scientists and everyday folk disagree. With respect to 'race', Glasgow defends the authority of common sense and argues that scientists are not the arbiters of the meaning of 'race'. Andreasen [2000; 2004; 2005] argues, on the other hand, that one need not reject the cladistic concept because it deviates somewhat form common sense. She uses the causal theory of reference as well as examples from the history of science to support her point. Yet she also raises the possibility that there might be two somewhat divergent meanings of 'race' — one folk, the other scientific — that coexist in relative harmony.

6 CONCLUSION

When addressing questions about the biological reality, or lack thereof, of race, it is often assumed that scientists have shown, conclusively, that races are biologically unreal. I hope to have shown, however, that the answer to this question is not as straightforward as many make it seem. Not only are many of the arguments against the biological reality of human race problematic, there are a number of newer biological conceptions that are not addressed by the standard arguments against the biological reality of race. Of these, phylogenetic conceptions appear to be the most promising.

[34] See Atran [1990] and Dupré [1993] for a similar point about other types of biological categories.

BIBLIOGRAPHY

[American Anthropological Association, 1998] American Anthropological Association. Statement on race. *American Journal of Physical Anthropology*, 101: 569–570, 1998.

[Andreasen, 1998] R. Andreasen. A new perspective on the race debate. *British Journal for the Philosophy of Science*, 49: 199–225, 1998.

[Andreasen, 2000] R. Andreasen. Race: biological reality or social construct? *Philosophy of Science*, 67 (Proceedings): 653–666, 2000.

[Andreasen, 2004] R. Andreasen. The cladistic race concept: a defense. *Biology and Philosophy*, 19(3): 425–442, 2004.

[Andreasen, 2005] R. Andreasen. The meaning of 'race': folk conceptions and the new biology of race. *Journal of Philosophy*, Vol. CII (2): 94–106, 2005.

[Appiah, 1996] K. A. Appiah. Race, culture, identity: misunderstood connections. In K. A. Appiah and A. Gutmann (eds.), *Color Conscious: The Political Morality of Race*. Princeton, N. J.: Princeton University Press, 30–105, 1996.

[Atran, 1990] S. Atran. *Cognitive Foundations of Natural History*. Cambridge, MA: Cambridge University Press, 1990.

[Bamshad and Olson, 2003] M. Bamshad and S. Olson. Do Races Exist? *Scientific American*, November 10, 2003.

[Banton and Harwood, 1975] M. Banton and J. Harwood. *The Race Concept*. London: David and Charles, 1975.

[Boyd, 1991] R. Boyd. Realism, anti-foundationalism, and the enthusiasm for natural kinds. *Philosophical Studies*, 61: 127–148, 1991.

[Cavalli-Sforza *et al.*, 1994] L. L. Cavalli-Sforza *et al.*. *The History and Geography of Human Genes*. Princeton, N. J.: Princeton University Press, 1994.

[Coon, 1963] C. S. Coon. *The Origin of Races*. New York, N. Y.: Alfred A. Knopf, 1963.

[Cuvier, 1812] G. Cuvier. *Recherches sur les Ossemens Fossiles*. Paris: Deterville, 1812.

[Diamond, 1994] J. Diamond. Race without color. *Discover*, November, 83–89, 1994.

[Dobzhansky, 1953] T. Dobzhansky. *Genetics and the Origin of Species*, 3^{rd} ed., New York: Columbia University Press, 1953.

[Dobzhansky, 1955] T. Dobzhansky. *Evolution, Genetics, Man*, New York: Wiley and Sons, 1955.

[Dupré, 1993] J. Dupré. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press, 1993.

[Edwards, 2003] A. W. F. Edwards. Human Genetic Diversity: Lewontin's Fallacy. *BioEssays*, 25: 798–801, 2003.

[Ereshefsky, 2001] M. Ereshefsky. *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge, MA: Cambridge University Press, 2001.

[Futuyma, 1986] D. Futuyma. *Evolutionary Biology*, 2^{nd} Edition, Sunderland, MA: Sinauer Press, 1986.

[Gannett, 2004] L. Gannett. The biological reification of race. *British Journal for the Philosophy of Science*, 55: 323–345, 2004.

[Garn, 1961] S. M. Garn. *Human Races*. Springfield, IL: Thomas, 1961.

[Ghiselin, 1974] M. Ghiselin. A radical solution to the species problem. *Systematic Zoology*, 23: 536–544, 1974.

[Glasgow, 2003] J. Glasgow. On the new biology of race. *The Journal of Philosophy*, Vol. C (9): 456–74, 2003.

[Gobineau, 1853-54] J. A. Gobineau. *Essai sur L'Inégalité des Races Humaines*. London: Heinemann, 1853-54.

[Goodman, 1978] N. Goodman. *Ways of Worldmaking*. Indianapolis, IN: Hackett Publishing Company, 1978.

[Gould, 1977] S. J. Gould. *Ever Since Darwin: Reflections in Natural History*. New York, N. Y.: W. W. Norton, 1977.

[Gould, 1981] S. J. Gould. *The Mismeasure of Man*. New York, N. Y.: W. W. Norton & Co., 1981.

[Hardimon, 2003] M. Hardimon. *Journal of Philosophy*, Vol. C (9): 437–455, 2003.

[Haslanger, 2005] S. Haslanger. What are we talking about? The semantics and politics of social kinds. *Hypatia*, forthcoming, 2005.

[Hennig, 1966] W. Hennig. *Phylogenetic Systematics*. Urbana, IL: University of Illinois Press, 1966.

[Hirschfeld, 1996] L. A. Hirschfeld. *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds*. Cambridge, MA: The MIT Press, 1996.

[Hirschman, 2000] Hirschman *et al.*. The meaning and measurement of race in the U.S. census. *Demography*, XXXVII: 381–93, 2000.

[Hull, 1965] D. L. Hull. The effect of essentialism on taxonomy: two thousand years of stasis. *British Journal for the Philosophy of Science*, 15: 314–26; 16: 1–18, 1965.

[Hull, 1970] D. L. Hull. Contemporary systematic philosophies. *Annual Review of Ecology and Systemtics*, 1: 19–53, 1970.

[Hull, 1976] D. L. Hull. Are species really individuals? *Systematic Zoology*, 25: 174–191, 1976.

[Hull, 1978] D. L. Hull. A matter of individuality. *Philosophy of Science*, 45: 335–360, 1978.

[Hull, 1998] D. L. Hull. Species, subspecies, and races. *Social Research*, 65 (2): 351–367, 1998.

[Keita and Kittles, 1997] S. O. Y. Keita and R. A. Kittles. The persistence of racial thinking and the myth of racial divergence. *American Anthropologist*, 99(3): 534–544, 1997.

[Keita *et al.*, 2004] S. O. Y. Keita *et al.*. Conceptualizing human variation. *Nature Genetics Supplement*, 36(11): S17–27, 2004.

[King and Stansfield, 1990] R. C. King and W. D. Stansfield. *A Dictionary of Genetics*, New York: Oxford University Press, 1990.

[Kitcher, 1999] P. K. Kitcher. Race, ethnicity, biology, culture. In L. Harris (ed.), *Racism*, Amherst, N.Y.: Humanity Books, 1999. Reprinted in P. Kitcher (ed.), *In Mendel's Mirror*, Cambridge, MA: Oxford University Press, 230–257, 2003.

[Legge *et al.*, 1996] J. Legge *et al.*. Genetic criteria for establishing evolutionarily significant units in Cryans buckmoth. *Conservation Biology*, 10: 85–98, 1996.

[Lewontin, 1972] Lewontin, R. C. The apportionment of human diversity. *Evolutionary Biology*, 6: 381–392, 1972.

[Lewontin *et al.*, 1984] R. C. Lewontin *et al.*. I. Q.: the rank ordering of the world. In *Not in Our Genes: Biology, Ideology, and Human Nature*. New York, N. Y.: Pantheon Books, 83–129, 1984.

[Livingstone, 1964] F. B. Livingstone. On the nonexistence of human races. In A. Montagu (ed.), *The Concept of Race*. New York, N. Y.: Collier Books, 46–60, 1964.

[Longino, 1990] H. E. Longino. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press, 1990.

[Machery and Faucher, forthcoming] E. Machery and L. Faucher. Why do we think racially? Culture, evolution, and cognition. In H. Cohen and C. Lefebvre (eds.), *Handbook of Categorization in Cognitive Science*, Elsevier, forthcoming.

[Mallon, forthcoming] R. Mallon. 'Race': normative, not metaphysical or semantic. *Ethics*, forthcoming.

[Mayr, 1942] E. Mayr. *Systematics and the Origin of Species*, New York: Columbia University Press, 1942.

[Mayr, 1959] E. Mayr. Typological versus population thinking (1959). In E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, A Bradford Book, 157–160, 1994.

[Miththapala, 1996] Miththapala *et al.*. Phylogeographic subspecies recognition in leopards (*Panthera pardus*) – Molecular Genetic Variation. *Conservation Biology*, 10: 1115–1132, 1996.

[Montagu, 1941] A. Montagu. The concept of race in the human species in the light of genetics (1941). In A. Montagu (ed.), *The Concept of Race*. New York, N.Y.: Collier Books, 1–11, 1964.

[Morton, 1839] S. G. Morton. *Crania Americana*. Philadelphia, PA: John Penington, 1839.

[Morton, 1844] S. G. Morton. Observations on Egyptian ethnography, derived from anatomy, history, and the monuments. *Transactions of the American Philosophical Society*, 9: 93–159, 1844.

[Morton, 1849] S. G. Morton. Observations on the size of the brain in various races and families of man. *Proceedings of the Academy of Natural Sciences, Philadelphia*, 4: 221–224, 1849.

[Mountain, 1997] Mountain *et al.*. Multilocus genotypes, a tree of individuals, and human evolutionary history. *American Journal of Human Genetics*, 61: 705–718, 1997.

[Nei and Roychoudhury, 1972] M. Nei and A. K. Roychoudhury. Gene differences between Caucasian, Negro, and Japanese populations. *Science*, 177: 434–435, 1972.

[Nei and Roychoudhury, 1974] M. Nei and A. K. Roychoudhury. Genetic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids. *American Journal of Human Genetics*, 26: 421–443, 1974.
[Nei and Roychoudhury, 1982] M. Nei and A. K. Roychoudhury. Genetic relationships and evolution of human races. *Evolutionary Biology*, 14: 1–59, 1982.
[Nei and Roychoudhury, 1993] M. Nei and A. K. Roychoudhury. Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution*, 10(5): 927–943, 1993.
[Nott and Gliddon, 1854] J. C. Nott and G. R. Gliddon. *Types of Mankind*. Philadelphia, PA: Lippincott, Grambo & Co., 1854
[Nott and Gliddon, 1868] J. C. Nott and G. R. Gliddon. *Indigenous Races of the Earth*. Philadelphia, PA: J. B. Lippincott, 1868.
[Omi and Winant, 1994] M. Omi and H. Winant. *Racial Formation in the United States: From the 1960s to the 1990s*. New York, N. Y.: Routledge, 2nd edition, 1994.
[Pigliucci and Kaplan, 2003] M. Pigliucci and J. Kaplan. Biological race and its applicability to humans. *Philosophy of Science*, 70: 1161–1172, 2003.
[Prichard *et al.*, 2000] J. K. Prichard *et al.*. Inference of population structure using multilocus genotype data. *Genetics*, 155: 945–959, 2000.
[Ridley, 1986] M. Ridley. *Evolution and Classification: The Reformation of Cladism*. London: Longman Scientific and Technical, 1986.
[Ridley, 1993] M. Ridley. *Evolution*. Boston, MA: Blackwell Scientific Publications, 1993.
[Risch, 2002] N. Risch, *et al.*. Categorization of humans in biomedical research. *Genome Biology*, 3(7): 1–12, 2002.
[Root, 2000] M. Root. How we divide the world. *Philosophy of Science*, 67: 628–639, 2000.
[Root, 2001] M. Root. The problem of race in medicine. *Philosophy of the Social Sciences*, 31(1): 20–39, 2001.
[Root, 2003] M. Root. The use of race in medicine as a proxy for genetic differences. *Philosophy of Science*, 70: 1173–1183, 2003.
[Rosenberg, 2002] N. Rosenberg *et al.*. Genetic structure of human populations. *Science*, CCXCVIII, 2381–85, 2002.
[Sesardic, 2003] N. Sesardic. Book review of Naomi Zack's *Philosophy of Science and Race*. *Philosophy of Science*, 70: 447–449, 2003.
[Shaffer and McKnight, 1996] H. Shaffer and M. McKnight. The polytypic species revisited. *Evolution*, 50: 417–433, 1996.
[Smedley, 1993] A. Smedley. *Race in North America: Origin and Evolution of a Worldview*. Boulder, CO: Westview Press, 1993.
[Sober, 1980] E. Sober. Evolution, population thinking, and essentialism. In E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 161–189, 1980.
[Sober, 1993] E. Sober. *Philosophy of Biology*. Boulder, CO: Westview Press, 1993.
[Stone *et al.*, 2001] G. Stone *et al.*. Differential success in northwards range expansion between ecotypes of the marble gallwasp *Andricus kollari*: a tale of two lifecycles. *Molecular Ecology*, 10: 761–778, 2001.
[Sundstrom, 2002] R. Sundstrom. Race as a human kind. *Philosophy and Social Criticism*, 28(1): 91–115, 2002.
[Templeton, 1999] A. R. Templeton. Human races: a genetic and evolutionary perspective. *American Anthropologist*, 100(3): 632–650, 1999.
[Thorne and Wolpoff, 1992] A. G. Thorne and M. H. Wolpoff. The multiregional evolution of humans. *Scientific American*, Apr.: 76–83, 1992.
[Vigilant, 1991] L. Vigilant *et al.*. African populations and the evolution of human mitochondrial DNA. *Science*, 253: 1503–1507, 1991.
[Vincente, 2001] J. G.Vincente *et al.*. Identification and origin of *Xanthomonas campestris pv. campestris* races and relate pathovars. *Phytopathology*, 91: 492–499, 2001.
[Wilson and Cann, 1992] A. C. Wilson and R. L. Cann. The recent African genesis of humans. *Scientific American*, 266: 68–73, 1992.
[Wilson and Brown, 1953] E. O. Wilson and W. L. Brown. The subspecies concept and its taxonomic application. *Systematic Zoology*, 2: 97–111, 1953.
[Wolpoff, 1989] M. H. Wolpoff. Multiregional evolution: the fossil alternative to Eden. In P. Mellars and C. B. Stringer (eds.), *The Human Revolution: Behavioral and Biological Perspectives on the Origins of Modern Humans*. Edinburgh: Edinburgh University Press, 62–108, 1989.

[Wolpoff *et al.*, 1988] M. H. Wolpoff *et al.*. Modern human origins. *Science*, 241: 772–773, 1988.
[Wright, 1994] L. Wright. "One Drop of Blood',' *The New Yorker*, July 25, 1994, 46-55, 1994.
[Wright, 1995] L. Wright. Who's black, who's white, and who cares. *Vanderbilt Law Review*, XLVIII: 513–569, 1995.
[Zack, 1997] N. Zack. Race and philosophic meaning. In N. Zack (ed.), *Race/Sex: Their Sameness, Difference, and Interplay*. New York, N. Y.: Routledge, 1997.
[Zack, 2002] N. Zack. *Philosophy of Science and Race*. New York: Routledge, 2002.

This page intentionally left blank

Part V
Special Topics

This page intentionally left blank

FORMALISATIONS OF EVOLUTIONARY BIOLOGY

Paul Thompson

. . . it is [mathematics] which reveals every genuine truth, for it knows every hidden secret, and bears the key to every subtly of letters; whoever, then, has the effrontery to study physics while neglecting mathematics, should know from the start that he will never make his entry through the portals of wisdom

(Thomas Bradwardine *Tractatus de Continuo*,[1] *circa* 1330s)

Philosophy is written in this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.

(Galileo Galilei, *Il saggiatori*,[2]1623)

INTRODUCTION

A formalisation of a theory is an abstract representation of the theory expressed in a formal deductive framework[3] for which there is a complete specification of:

1. what constitutes a well-formed-formula (wff), and
2. all the permissible rules of inference.

[1]As quoted by J.A. Weisheipl, "Galileo and his precursors," in E. McMullin (ed.) *Galileo: Man of Science*.

[2]Commonly known in English as *The Assayer* (Middle English "assay" assimilated to French "essayer"). A superb translation by Stilman Drake can be found in: C.D. O'Malley and S. Drake (eds.) *Controversy on the Comets of 1618*, Philadelphia, 1960.

[3]This is a broader conception than that employed by logicians. The narrower concept conceives of a formal system in terms of first-order predicate logic with identity.

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

Consequently, a theory can be formalised in first-order predicate logic with identity, in set theory, in matrix algebra and, indeed, any branch of mathematics satisfying (1) and (2).

A theory (for example Euclidean geometry or arithmetic) also requires:

3. a complete specification of a set of primitive formulae of the theory (those which cannot be inferred from any others and from which all other formulae of the theory will be inferred.[4])

A *scientific* theory also has to include:

4. an interpretation (a semantics) for the formal calculus (the syntax).

How the semantics is provided differs in different conceptions of the formalisation of scientific theories.

1 THE PATH TO A GALILEAN CONCEPTION OF SCIENTIFIC THEORIES

In the early part of the 20^{th}-century, formalisation in science was virtually synonymous with the representation of a theory in first-order predicate logic with identity. This was a result of the dominance through to the 1950's of logical positivism and its descendent logical empiricism in philosophy of science. As noted below, this view has various names of which "the syntactic conception" is the most descriptive. The logical empiricist conception provided the semantics through correspondence rules. Through the 1950's and into the1960's, serious logical and conceptual difficulties with the logical empiricist conception were forcefully articulated. In the forefront of this wave of critical material was Willard van Orman Quine's, "Two Dogmas of Empiricism" [Quine, 1951]. One of the most influential criticisms was Thomas Kuhn's, *The Structure of Scientific Revolutions* [Kuhn, 1962]. A common thread in much of the criticism was the failure of logical empiricism to capture the holist nature of scientific theories. Much of the blame for this rested with the employment of correspondence rules to provide a semantics for the theory.

In the latter part of the 1960's, an alternative conception was developed in which theories were represented in set theory or in a state space. This conception was widely called "the semantic conception." It builds on earlier work on model-theoretic semantics of formal systems (e.g., Alfred Tarski and Evert Beth). This conception treats scientific theories holistically, avoiding, thereby, one of the central defects of the logical empiricist conception of theories. It is important to be clear that a syntactic conception of theories is not wedded to providing semantics by means of correspondence rules (about which I shall have more to say later). Indeed, following Taski, one could provide a model-theoretic semantics and, thereby, capture the holistic character of theories. Those who espouse the semantic conception accept that this provides a more robust syntactic conception of theories, but

[4]Kurt Gödel established that formal systems cannot be both complete and consistent.

claim, correctly I think, that providing semantics via models makes the syntactic formalisation otiose.

The syntactic conception was forged within the framework of logical positivism/empiricism and was deeply informed by the rise of symbolic logic and the conviction that it was the appropriate formal calculus for the formalisation of scientific theories. The semantic conception was fashioned against this background, denying that first-order predicate logic with identity was the most appropriate formal calculus and denying that correspondence rules provided the required semantics. The label "semantic conception" captures the rejection of the logical empiricist method of providing a semantics but it does not reflect fully the rejection of first-order predicate logic as the appropriate formal calculus. Once this constraint on the appropriate formal system is relaxed, the formal calculus of any domain of mathematics can be used to provide a formalisation of a theory. This is entirely consistent with the semantic conception but is not fully reflected in the name and the emphasis of its early advocates on set theory and state-spaces obscures the richer mathematical resources available. Hence, I suggest dubbing a mature conception of the formalisation of scientific theories "the Galilean conception," echoing his emphasis on mathematics as the revealer of genuine scientific truth. As the opening quotation establishes, Bradwardine takes temporal precedence in advocating that mathematics is the key to understanding nature but Galileo, in his famous quotation, is more explicit about the role of mathematics as the language of science. The Galilean conception explicitly expands the scope of the mathematical domains that are employed to formalise a theory to encompass all of mathematics. This is not in conflict with the semantic conception; it is an extension of it. The Galilean conception contends that the tacit limitation on a formalisation to set theory or state spaces, as contained within central expositions of the semantic conception, is an unintended consequence of the framework (logical empiricism) with which the conception was being contrasted during its development.[5]

This explicit expansion of the resources available is especially important in light of recent biological theorising with respect to (non-linear dynamical systems (e.g., neurobiology, population dynamics).[6] However, as we shall see below, the use of the rich array of mathematical resources has a prior history even in evolutionary biology. Population genetics from Fisher, Haldane and Wright onward has provided a formalisation of population dynamics using probability theory and statistics.

Each of these conceptions can be viewed as a temporal stage in the development of a rich conception of the formalisation of scientific theories. Each stage is a refinement of the previous stage; the different conceptions are less in conflict than they are points in an intellectual journal. In what follows, I will describe

[5]A careful reading of F. Suppe, P. Suppes and B. van Fraassen, the key developers and promoters of the semantic conception, makes clear that they understood "models of theory" in a rich way that was not at all restricted to set-theoretical or state space presentations.

[6]See, for example, [Segal, 1989; Amit, 1989; Maurer, 1999; Murray, 1993; Renshaw, 1991].

these conceptions in more detail and explore the relationship of each to biological theorising.

1.1 The First-order Predicate Logic (Syntactic) Conception of Theories

This conception is a syntactic conception. That is, a scientific theory is formalised in terms of well-formed-formulae (wff's) of the syntax of first-order predicate logic with identity.[7] It is also an axiomatic-deductive structure. Some sets of wwf's capture those claims of the theory which cannot be deduced from other claims and from which all the other claims within the scope of theory can be deduced. This set constitutes the axioms of the theory. All other wff's of the theory are deducible from the axioms. Hence, for obvious reasons, this conception has also been called the axiomatic-deductive conception.[8] However, although it is indeed *an* axiomatic-deductive conception, this characterisation does not distinguish it from the conception of theories put forward by Patrick Suppes. On his conception of theories, which we will explore later, theories are formalised in set theory, and the axioms of a theory are specified in terms of a set-theoretical predicate. All other claims of the theory are expressed in set-theoretical terms and are a logical consequence of the axioms. Therefore, characterising a conception as axiomatic-deductive does not provide a unique descriptor. On the other hand, characterising a conception as syntactic does distinguish it from the other conceptions that have been advanced. Hence, "syntactic" is a more useful characterisation of the view set out in this section. This view held such sway in the first half of the 20^{th}-century that Hilary Putnam in 1962 dubbed it "The Received View."

The syntactic account arose within the school of philosophy known as logical positivism and in the wake of the publication of Bertrand Russell and Alfred North Whitehead's *Principia Mathematica*. *Principia Mathematica* codified mathematical logic and made possible the symbolic representation of ordinary language statements; those statements being rendered as well-formed-formulae in first-order predicate logic with identity. It also made possible the exploration of the logical (deductive) relationships among those formulae and, hence, among the linguistic statements they represent. The syntactic account's most influential and important formulation was worked out by the logical empiricists,[9] Rudolf Carnap, Carl Hempel, Ernest Nagel and Richard Braithwaite (see [Suppe, 1977] for an excellent exposition and historical account of this conception).

[7] First-order logics are those logics whose entities are individuals. In the case of first-order predicate logic, the objects are individual things. Subjects and predicates are symbolically represented and the quantifiers "all" and "some" are employed. Second-order logics are those logics whose entities are sets of individuals.

[8] Alfred Tarski provides a clear and influential account of the deductive method and its application in constructing mathematical theories in Tarski [1946].

[9] Logical empiricism is a descendent of logical positivism (see Giere and Richardson [1996], and Misak [1995]). Carnap proposed changing the name to reflect both its empiricist stance and the softening of some of the positions found in logical positivism.

According to this conception of theories, at the core of a theory are a few axioms (fundamental laws of that domain of science). The axioms can be expressed in ordinary language but their formal and definitive expression is in the language of symbolic logic. The axioms are laws of the highest generality within the theory. Ideally, they constitute a consistent set, no one of which can be derived from any subset of the others. From these axioms in their symbolic-logic formulation all the rest of the formula (laws) of the domain of science, in principle, can be deduced.[10] These deductions employ the inferential machinery of first-order predicate logic. Invariably, in the case of scientific theories, numerous subsidiary assumptions must be added in any deduction in order to derive a law from the axioms; it is held that perfect knowledge, although unattainable, would make these assumptions unnecessary.

Just as the terms used in ordinary language need to be given meaning (to be defined) in order for a grammatically correct sentence to be understood, so too do the terms in a well-formed formula in symbolic logic need to be given meaning (be defined). In the logical empiricist formulation, this deductively related set of statements is given empirical meaning by definitions — called correspondence rules — which *ultimately* link theoretical terms (e.g., population, fertile, disease, motility, polymorphonuclear, chemotaxis, gene, and so forth) to observations (e.g., a theoretical term like "fertile" is partially defined by reference to the outcomes of numerous sexual events of a specified kind under specified conditions). Some theoretical terms are defined by reference to one or more other theoretical terms. Ultimately, any chain of such definitions must end in theoretical terms that are defined by reference to observations. In this way, the theory as a whole is given empirical meaning. Because of this complex interconnection of theoretical terms, the meaning of any one term is seldom independent of the meaning of many if not all of the other terms of the theory. Hence, theories have a global meaning structure: changes to the meaning of one term will have consequences for the meaning of many and usually all the other terms of the theory.

The influence of this conception on the discussions of theory structure in biology was profound. The earliest impact was on theoretically engaged biologists. For example, very early in its development, J.H. Woodger [1937; 1939] argued for an application of the axiomatic method in biology and provided his own axiomatic account of selected biological theories. C.H. Waddington [1968–72] also promoted

[10]Two important features of an entirely satisfactory formal system (and, hence, a scientific theory) are completeness and consistency. A formal system is complete if, within its domain, all the possible formulae *or* their negations are derivable (provable). A theory is consistent if there is no formula *and* its negation that can both be derived. Unfortunately, Kurt Gödel, in a now famous proof, established that no formal system can be both complete (every formula or its negation being provable) and consistent (no formula and its negation being provable) (a translation of the original 1931 paper can be found in [Van Heijenoort, 1966], and an excellent exposition of the proof is in [Nagel and Newman, 1958]). Gödel's proof was directed at the problem of the consistency of arithmetic but it applies to all formal systems. This feature of formal systems is not one that for most purposes outside mathematics matters profoundly. If either completeness or consistency has to be sacrificed in order to employ a formal system, it is usually advantageous to relax the requirement of completeness.

and employed the logical empiricist view of theories. Among philosophers, opinion on the applicability of this conception to biology was divided. Morton Beckner [1959] proved to be prescient in his account of biological theories as a family of models. He did not embrace the logical empiricist account assuming that its application to biology was at best limited. Thomas Goudge [1961] was quite clear, whatever applicability the conception might have in physics, it did not capture important features of biological theories and explanation. Two philosophers who, during the 1970's and 1980's stood squarely behind the applicability of this conception to biology are Alexander Rosenberg [1985a] and Michael Ruse [1973]. Others with varying degrees of qualification and subtlety have argued in favour of this conception's relevance to biological theorising (e.g., David Hull [1974])

The account that Ruse provides of a biological theory that can be given a syntactic conception formalisation focuses on population genetics [Ruse, 1973, pp. 32–37]. The axioms of population genetics are Mendel's Laws: the law of segregation and the law of independent assortment. The thrust of Ruse's argument is that the Hardy–Weinberg Equilibrium, a fundamental principle of population genetics, can be derived from Mendel's second law (the law of independent assortment). What Ruse provides is a sketch of an axiomatic-deductive framework. He does not cast Mendel's laws as formulae in first-order predicate logic. He does not include other required axioms: the axiom of differential reproduction, the axiom of linkage, to mention only two key ones. One derivation is a little light. But perhaps the most telling feature of Ruse's argument is that his structure for Mendel's second law and the deductive machinery he employs is not first-order predicate logic but that of matrix algebra — albeit a very elementary application of matrix algebra. Hence, his account seems more like an instance of what I later in this paper introduce as the Galilean conception of scientific theories.

The account that Alexander Rosenberg [1985a] provides builds heavily on an axiomatisation of selection theory constructed by Mary Williams [1970]. Williams' axiomatisation of selection theory is a superb and sophisticated example of the application of an axiomatic method in biology. It is not surprising, therefore, that Rosenberg makes it the centrepiece of his argument in favour of the logical empiricist conception. Although Williams uses "evolution" in the title of the paper, the axiomatisation is, in fact, of selection theory. The axiomatisation makes scant reference to genetics. Rosenberg attempts to mitigate this difficulty by arguing that evolutionary theory ***is*** selection theory — genetics (heredity) is just a background condition.[11] I have argued elsewhere why this move fails to rescue this axiomatisation from the problem of its isolation from evolutionary theory as a whole [Thompson, 1983]. In Part II below, I argue that evolutionary theory is a composite of selection theory, genetics, ecology and a host of other domains. Within this conception of evolutionary theory, Williams' axiomatisation is powerful and

[11]This stands in stark contrast to Ronald A. Fisher's opening sentence in the Preface to his classic work *The Genetical Theory of Natural Selection* [1930]. Fisher writes, "Natural Selection is not evolution."

important.[12] As an example of a syntactic formalisation of evolutionary theory it fails. Indeed, in addition to axiomatising only one part of a robust theory of evolution, her axiomatisation draws heavily on set theory, not first-order predicate logic. In the next section, it will become clear that set-theoretical axiomatisations fall within the semantic conception of theories and more broadly into the Galilean conception of theories.[13] In addition, in Part II, it will become clear that Williams' axiomatisation as a set-theoretical axiomatisation is an important contribution to the formalisation of a robust evolutionary theory.

By the late 1970's, logical empiricism had declined in influence. The early criticisms of Quine and later of Kuhn had taken a significant toll on the credibility of the programme. But the difficulties continued to emerge. Correspondence rules, for example, were totally inadequate to the task of providing a semantics for scientific theories. The most insightful, clear and devastating critique of correspondence rules was provided by Kenneth Schaffner [1969]. He highlighted the fact that relating a theory to phenomena requires the employment of laws from other independent theories, a fact that correspondence rules could not accommodate.

Although employing correspondence rules was the logical empiricist's dominant method of providing a semantics for well-formed formulae and, more broadly, syntactic systems, there is a large body of literature exploring a variety of ways in which the semantics can be provided. In the context of scientific theories, all are aimed at connecting the terms in the symbolic structure to phenomena (to the empirical world).

An alternative proposed by Alfred Tarski [1944; 1953; 1956] employs models. For Tarski, a "model" designates a structure through which the abstract symbolic formulae (wff's) of the theory can be given meaning, "A possible realization in which all valid sentences of a theory T are satisfied[14] is called a *model* of T." [1953, p. 11]. This is a robust and important method of providing a meaning structure for a theory formalised in first-order predicate logic and is a stepping-stone to the semantic account of theories. The model is a "true" interpretation of axiomatic-deductive systems. It is the fact that the model provides a *true interpretation* of the set of symbolic formulae that qualifies it as a meaning-structure for that

[12] Rosenberg and Williams published two joint papers [1985b; 1986] which continue the perspective Rosenberg took in his 1985 book [1985a]. One reasonably might assume from these joint ventures that Williams endorses Rosenberg's interpretation of her work. However, Williams' contribution can, and I claim should, be severed from this interpretation. Hers is not a first-order predicate logic axiomatisation interpreted via correspondence rules. She has provided an archetypal set-theoretic axiomatisation and it is better interpreted as providing a central component of a semantic or Galilean conception of the formalisation of evolutionary theory.

[13] The subtitle of Williams' paper, "A Mathematical Model," correctly suggests that she is constructing a mathematical model, not a first-order predicate logic axiomatisation. As a mathematical model it falls within a Galilean conception of theories.

[14] "Satisfied" means "rendered true." Hence, a set of expressions in a formal syntax is satisfied if all the expressions are true and, hence, consistent. Hence, for Tarski a model for the syntactic formalisation of a theory T is a realisation (an interpretation) of T in which all the expressions of T are true. If T is an axiomatic-deductive structure, an interpretation of the axioms of T that renders them all true constitutes a model of T. All other expressions are a deductive consequence of the axioms and, hence, will be true in that interpretation.

set of formulae. The force of the requirement that the interpretation be "true" is to demand that the model "satisfy" "all" the formula in the set. A formula is satisfied if the interpretation given by the model yields a specification that is well founded, that is, understood and widely believed by those in a position to judge. In the case of a scientific theory, the resulting specification must be empirically acceptable. That is, it is consistent with the phenomena within its scope.

Consider a simple example of this concept of satisfaction. The formula:

$$(x)(y)(((x = J)\&(y = B)) \rightarrow ((T_{xy} \rightarrow L_{yx})\&(T_{yx} \rightarrow L_{xy})))$$

is satisfied by the statement:

If John speaks to Bill then Bill listens carefully to John and if Bill speaks to John then John listens carefully to Bill.

This is a simplistic example since it provides an *ordinary-language* interpretation of a *single* formula. Model interpretations of scientific theories are most often formulated in a mathematical language. In addition, since a theory T will almost always be an interconnected set of formulae, the model of T must simultaneously satisfy *all* the formulae of the theory.

Of crucial importance to the use of models for interpreting formal systems is the fact that more than one interpretation can be given for a formula or set of formulae. Game theory, for example, has been successfully employed in ecology, economics, and international relations. The abstract structure of game theory can be expressed as an abstract formal system. In each of the discipline-specific applications, different models of the abstract system are being provided. In each case, the meanings of the terms as well as the empirical truth of the interpreted formulas can be different. And, in each case, the model satisfies (provides a semantics for) the abstract formal system if the resulting expressions are empirically acceptable.

This method of providing an interpretation for an abstract formal system turns out to be extremely powerful in the sciences and in mathematics. One of the secondary effects of its development was to call into question the utility of abstract formal structures in the context of scientific theories. Several features of scientific theories suggested that the syntactic formalisation was unachievable and, more importantly, unnecessary. A few of the features are: that anything approaching a full formal account of any actual scientific theory has proved elusive, correspondence rules, as indicated, as a method of interpretation have significant problems and almost all actual scientific theories are model-theoretic structures. As a result, some philosophers of science began to question the added value of a formalisation in first-order predicate logic given that the appropriate models that interpret the formal system could be specified directly. It was considered unimportant whether one could specify the formal system that a given model satisfied, why would one want to? As long as the model can be adequately specified without reference to the formal system and as long as the model can be used to achieve everything for which one might appeal to scientific theory, there is no added value. The semantic account of theories emerged from this view that scientific theories could be construed directly as models.

1.2 The Semantic Conception of Scientific Theories

The semantic conception is so called because scientific theories are formalized in terms of models (semantic structures) and, hence, an adequate formal approach to the structure of scientific theories consists in the direct specification of models and not in the specification of a first-order predicate logic formulation (a syntax). On the "received view" version of the syntactic conception, the semantics of a theory are provided by correspondence rules. As indicated above, one could provide the semantics, following Tarski, model-theoretically.

Briefly stated, with a more detailed exposition to follow, on a semantic conception, the semantics of a theory are provided directly by defining a class of models. For Patrick Suppes, the class of models is directly defined by specifying a set-theoretical predicate. For Bas van Fraassen and Frederick Suppe, the class of models is defined in terms of a phase space or state space (i.e., topologically). One point of difference between van Fraassen and Suppe is that van Fraassen identifies theories with state spaces whereas Suppe understands state spaces as "canonical iconic models of theories" [Suppe, 1972a, p. 161, note 18] or "canonical mathematical replicas of theories" [Suppe, 1977, pp. 227–228, note 565].

A broad outline of the points of agreement and the points of difference between the Received View and the semantic is easily provided. The three major components of a formalisation of a scientific theory on the Received View are: (1) the mathematical domain used to provide the syntax for the theory is first-order predicate logic, (2) a formalised theory is an axiomatisation and (3) the semantics of the theory are provided by correspondence rules. The semantic conception rejects the stricture of (1), accepts (2) if "axiom" is understood broadly enough to allow fundamental theorems of the theory to be classed as axioms when expressed in mathematical domains such as topology, set theory, and so on, and rejects (3) entirely.

Frederick Suppe has traced the origin of the semantic view of theories to John von Neumann in the 1940's [Suppe, 1989].[15]) Two other early initiators and advocates were Evert Beth [1948; 1949; see also 1961] and Patrick Suppes [1957] in his *Introduction to Logic*. Suppes further develops his account in several publications [Suppes, 1962; 1967; 1968]. Beth, and following him Bas van Fraassen [1970; 1972; 1980; 1981], advanced a state space approach while Suppes advanced a set-theoretical predicate approach. Fred Suppe has also been a developer and champion of the state space approach [1967; 1972; 1977; 1989]. As I will show below, Mendelian genetics can be formulated quite naturally on either approach.[16]

In 1957, Suppes suggested that scientific theories are more appropriately formalised as set-theoretical predicates. Shortly thereafter, in 1961, Robert Stoll in his *Set Theory and Logic* [Stoll, 1961] made a similar claim about the formaliza-

[15]Suppe [1989] provides an excellent history of the development of the semantic conception, and a defence of scientific realism within this conception.

[16]A set-theoretical version of the formulation of Mendelian genetics is as follows:

T: A system $\beta_{-} = \langle P, A, f, g\rangle$ is a Mendelian breeding system if and only if the following axioms are satisfied:

tion of informal theories, of which scientific theories are instances. Suppes wrote a number of papers during the 1960's in which he argued that scientific theories were better represented as a set-theoretical predicate (see, for example [Suppes, 1962]). In 1967, he wrote a brief non-technical account of his view [Suppes, 1967]. In this paper, he quite clearly sets out his reasons for rejecting what he calls the "standard sketch of scientific theories" and for adopting a semantic account. His central conclusion is that scientific theories are not appropriately or usefully formalised as axiomatisations in first-order predicate logic but rather in set theory.

The main thrust of his argument was that correspondence rules (he calls them co-ordinating definitions) "do not in the sense of modern logic provide an adequate semantics for the formal calculus" [Suppes 1967, p.57]. One should instead talk about models of the theory. These models are non-linguistic entities that are

---

Axiom 1: The sets P and A_{-} are finite and non-empty.
Axiom 2: For any $a \in P$ and $l, m_{-} \in A, f(a, I) \& f(a, m)$ iff $I = m$.
Axiom 3: For any $a, b, \in P$ and $I \in A, g(a, I) \& g(b, I)$ iff $a = b$.
Axiom 4: For any $a, b \in P$ and $l \in L$ such that $f(a, l)$ and $f(b, l), g(a, l)$ is independent of $g(b, l)$.
Axiom 5: For any $a, b \in P$ and $l, m \in L$ such that $f(a, l)$ and $f(b, m), g(a, l)$ is independent of $g(b, m)$.

Where P and A_{-} are sets and f and g are functions. P is the set of all alleles in the populations, A is the set of all loci in the population. If $a \in P$ and $I \in A$, then $f(a, l)$ is an assignment, in a diploid phase of a cell, of a to I (i.e., f is a function that assigns a as an alternative allele at locus l). If $a \in P$, and $I \in A$, then $g(a, I)$ is the gamete formed, by meiosis, with a being at I in the gamete (the haploid phase of the cell). Although more sophistication could be introduced into this example (to take account, for example, of meiotic drive, selection, linkage, crossing over, etc.), the example as it stands illustrates adequately the nature of a set-theoretical approach to the formalization of population genetic theory in its simple Mendelian system form.

Characterizing Mendelian genetics using a state space approach is more complicated. A theory on this view consists of the specification of a collection of mathematical entities (numbers, vectors, functions) used to represent states of the system within a state space (a topological structure), the behaviour of which is specified by three functions. These functions are commonly called laws but are not the same as laws in the received view. In the received view, laws are statements describing the behaviour of entities in the world. In the semantic view, laws are descriptions of the behaviour of mathematical systems of mathematical entities. The three laws are: laws of coexistence (which specify the physically possible set of states of the system); laws of succession (which specify the possible histories of the system); and laws of interaction (which specify the behaviour of the system under conditions of inputs from interaction with other systems). A theory also requires the specification of a set of measurable magnitudes (variables) (represented by a function defined on the state space). Statements which formulate propositions to the effect that a particular magnitude has a particular value at a particular time are elementary statements. A satisfaction function determines the sets of states which satisfy the assignment of a value to a physical magnitude.

For population genetic theory, the state space will be a Cartesian n-space where 'n' is a function of the number of possible pairs of alleles in the population. A law of coexistence to the effect that only alleles at the same locus can form pairs will select the class of physically possible pairs of alleles. States of the system (genotype frequencies of populations) are n-tuples of real numbers from zero to one and are represented in the state space as points. These are the measurable magnitudes. An example of a satisfaction function for the elementary statement "genotype Aa occurs with a frequency of 0.5" would specify the set of states in the state space that satisfy the statement. In this case, the set of states would be a Cartesian $(n - I)$-space, which is a subset of the state space. For population genetic theory, a central law of succession is the Hardy-Weinberg law.

highly abstract and are far removed from the empirical phenomena to which they will be applied. As I have argued [Thompson, 1983; 1985; 1986; 1987; 1988a; 1989], this feature is one of the major strengths of this conception since it accords with the actual scientific practice of using a multiple number of auxiliary theories (from mathematics and other domains of science) in the application of a particular scientific theory to phenomena.

Suppes suggests two reasons why the syntactic account is so widely and strongly held, despite what he argues are logical and practical weaknesses with it. First, philosopher's examples of scientific theories are usually fairly simple and, therefore, easily able to be given a linguistic formulation. Not surprisingly, most examples used to explicate and defend the syntactic account are drawn from Newtonian mechanics — and from a reasonably simple and sketchy account of it. Also not surprisingly, advocates of the semantic account discuss complex theories such as quantum mechanics (see [van Fraassen, 1972]), learning theory (see [Suppes, 1962]), and evolutionary theory (see [Beatty, 1980a; 1980b; Lloyd, 1983; 1984; 1986; 1987; Thompson, 1983b; 1985; 1986; 1987; 1988a]). Second, compared to the view that theories are axiomatic-deductive structures formulated in first-order predicate logic and partially interpreted by correspondence rules, far more mathematical sophistication is required to formulate a theory on the semantic account and to characterise explanation, prediction, and validity.

During the late 1960's and the 1970's the semantic account was consolidated and extended by a number of philosophers from a variety of perspectives (see, for example: [Suppe, 1967; 1972a; 1972b; 1974; 1976; van Fraassen, 1970; 1972; 1980; Sneed, 1971; Stegmuller, 1976]). Despite this coalescing of the account, there were, and continue to be, important differences of motivation and structure between the views of these advocates. Suppe, for example, is a scientific realist (see [Suppe, 1988]), whereas van Fraassen is a constructive empiricist (see, [van Fraassen, 1980]). During the 1980's John Beatty, Elisabeth Lloyd and I extended and applied the semantic account to biology and, in particular, evolutionary theory and genetics (see [Beatty, 1980a; 1980b; Lloyd, 1983; 1984; 1986; 1987; Thompson, 1983b; 1985; 1986; 1987; 1988a; 1988b; 1989]).[17]

One of the major features of theories on the semantic account is that the class of models, which specify directly in mathematical English the behaviour of a system, is an extra-linguistic, highly abstract entity which is most often quite removed from the phenomena to which it is intended to apply. For example, laws do not describe the behaviour of objects in the world, they specify the nature and behaviour of an abstract system.

The application of the model(s) to a particular empirical system requires the extra-theoretical assertion that the model(s) and the phenomena to which they are intended to apply are isomorphic (in algebraic contexts such as set theory, groups and rings) or homeomorphic (in topological contexts). This is an extra-theoretical

[17]Frederick Suppe applied the semantic account to aspects of taxonomy in the 1970a. But this seems to be an isolated application to biology prior to the early 1980s (see Suppe [1989]).

assertion because it not specified by the theory itself.[18]

Justifying the assertion that an isomorphism exists is a complex task which requires: the employment of a range of other scientific theories, the adoption of theories of methodology (e.g., theories of experimental design, goodness of fit, etc.), the employment of a variety of domains of mathematics (e.g., probability, statistics, linear algebra), etc. If a theory and the phenomena within its intended scope are shown to be isomorphic, then explaining and predicting outcomes within the model constitutes explaining and predicting outcomes in the empirical world. Advocates of the semantic view have seen the separation of the theory and the methods of its application and confirmation as a major logical, heuristic and methodological advantage of the conception.

An isomorphism or homeomorphism is a one-to-one correspondence between the elements of one or more sets resulting from a bijective mapping. A bijective mapping is a 'one-to-one' (injective) and 'onto' (surjective) mapping. A function f is a one-to-one mapping function if and only if:

$$(x_1 \neq x_2) \Leftrightarrow (f(x_1) \neq f(x_2)).$$

A function f is an onto mapping function if:

$$y \in Y \Rightarrow \exists x \in X : f(x) = y.$$

The essential feature of an isomorphism in the context of scientific theories is the assertion of the "sameness" of the structure and behaviour of the model and the empirical world (i.e., if there is a one-to-one, and onto, mapping which preserves relations, functions and constants). As a consequence of characterising the relation between a theory and the empirical world as an isomorphism or homeomorphism,

[18]In the syntactic conception, on the other hand, the semantics is provided by correspondence rules which *are* part of the theory and directly link the formal system to the phenomenal world. In effect the correspondence rules define an empirical model of the formal system. That empirical model is understood as logically equivalent to the phenomenal system to which the theory applies. It is for this reason that actual phenomena can be deduced from the statements of the theory. That is, the statements of the theory are laws that describe the *actual* behaviour of objects in the world. Hence, any behaviour deduced from the statements of the theory is either a prediction about what *actually* will happen under the specified circumstance or an explanation of what *actually* did happen under the specified circumstances. The interpreted formal system directly describes the behaviour of entities in the world. In a semantic conception, a theory is defined directly by specifying in mathematical English the behaviour of a system. Most importantly, laws do not describe the behaviour of objects in the world, they specify the nature and behaviour of an abstract system. This abstract system is, independently of its specification, claimed to be isomorphic to a particular empirical system. Establishing this isomorphism, as I shall argue below, requires the employment of a range of other scientific theories and the adoption of theories of methodology (e.g., theories of experimental design, goodness of fit, etc.). In essence, the substance of the difference between the two conceptions is that the semantic conception calls into question the possibility of providing an adequate semantics for a scientific theory by means of correspondence rules. And, calls into question the need for any reference to a formal system since the semantics can be provided directly by defining a mathematical model. Advocates of the semantic view have seen the separation of the theory and the method of its application as a major logical, heuristic and methodological advantage of the conception.

confirmation of a theory consists in establishing that the required isomorphism holds. Explanation and prediction consists in demonstrating that the element(s) in the model that bijectively maps to the empirical event being explained or predicted is a consequence of the structure and behaviour of the model, where the model is deemed to have an acceptable level of confirmation.[19]

The actual task of confirming a theory (model) is complex involving, among others, theories of measurement, experiment, experimental design, goodness of fit, data modelling, etc. A theory of measurement provides an agreed upon standard in terms of which observed phenomena are compared as well as a set of principles governing the conditions under which measurements are made. These principles ensure that the measurements are in accord with the theory of the experiment. A theory of the experiment specifies a broad conceptual framework within which experiments can take place. It specifies such things as what assumptions based on other scientific theories can be employed (for example electromagnetic theory and quantum theory when using an electron microscope in a biological experiment), the possibility and role of simplifying assumptions, correct patterns of inference, etc. A theory of experimental design specifies the exact nature of the technique of experimentation. The appropriate methods for controlling extraneous variables are an important component specified by a theory of experimental design. A theory of goodness of fit (often a statistical measure such as a X^2-test[20] (chi-squared test), an example of which is given in the population genetics section below[21]) specifies when normalised data (by application of a model of data, often a probabilistic or statistical theory, especially regression analysis[22] employing, for example, a

[19] A theory (mathematical model) explains, or predicts, a phenomenon, or number of phenomena, if:

- the system defined by the theory is isomorphic to the phenomenal system in which the set of phenomena to be explained occur, and
- the set of elements of the mathematical model which are mapped onto the set of relevant phenomenal objects within the phenomenal system can be shown, within the mathematical model, to be a consequence of the structure or behaviour of the model.

From an abstract mathematical point of view, explanation is a description of the structure and behaviour of phenomenal systems in terms of the structure and behaviour of a mathematical model. As such, the validity of the explanation is necessarily related to validity of the assertion of an isomorphic relationship between phenomenal system and mathematical model. This is similar to the standard view, in which the validity of an explanation is necessarily dependent upon the validity of the laws of the theory to which an appeal is being made. Not surprisingly, therefore, confirmation is central to the empirical enterprise in both cases. The assertion of an isomorphism, is the foundational link between a theory and a phenomenal system. Confirmation grounds the assertion. Confirmation is complex and multilayered.

[20] This test was devised by the British scientist, mathematician and philosopher of science, Karl Pearson, in 1900 (his positivist philosophy of science is largely contained in Pearson [1892], the chi-squared test is found Pearson [1990]). $X^2 = \sum$ (observed quantity — expected quantity)^2/(expected quantity). The quantity is often a frequency.

[21] Another technique for determining goodness of fit is to compare a multidimensional graphic representation of a theoretically predicted distributional landscape and distributional landscape of observed results. Hartl [2000, pp.88-94] provides an excellent example of this with respect to diffusion approximations of random genetic drift.

[22] For an excellent mathematical presentation of regression analysis see [Sen and Srivastava,

least squares method[23] sufficiently corresponds to the relevant elements of the theory. One of the most straightforward ways of assessing the correspondence is to represent the data in an 'observation space' analogous to the 'mathematical space' in which the theory is represented. In these spaces, the states of both the empirical system and the theories are points in the respective spaces. In the easiest case, namely that of a deterministic linear system, the comparison of the observation space and the phase space of the theory is uncomplicated and based on an identity relation — the observation space and the phase space will have the same dimensionality and the points representing states of the system will be identically located within the space.

There are two prominent versions of the semantic conception of theories: a set-theoretical version and a state space version.

An example of each can be given for Mendelian genetics. The set-theoretical version of the formulation of Mendelian genetics is:

T: A system $\beta = \langle P, A, f, g\rangle$ is a Mendelian breeding system if and only if the following axioms are satisfied:

Axiom 1: The sets P and A are finite and non empty.

Axiom 2: For any $a \in P$ and $l, m \in A$, $f(a, I)\&f(a, m)$ iff $I = m$.

Axiom 3: For any $a, b, \in P$ and $I \in A$, $g(a, I)\&g(b, I)$ iff $a = b$.

Axiom 4: For any $a, b \in P$ and $l \in L$ such that $f(a, l)$ and $f(b, l), g(a, l)$ is independent of $g(b, l)$.

Axiom 5: For any $a, b \in P$ and $l, m \in L$ such that $f(a, l)$ and $f(b, m), g(a, l)$ is independent of $g(b, m)$.

Where P and A are sets and f and g are functions. P is the set of all alleles in the populations, A is the set of all loci in the population. If $a \in P$ and $I \in A$, then $f(a, l)$ is an assignment, in a diploid phase of a cell, of a to I (i.e., f is a function that assigns a as an alternative allele at locus l). If $a \in P$, and $I \in A$, then $g(a, I)$ is the gamete formed, by meiosis, with a being at I in the gamete (the haploid phase of the cell). Although more sophistication could be introduced into this example (to take account, for example, of meiotic drive, selection, linkage, crossing over, etc.), the example as it stands illustrates adequately the nature of a set-theoretical approach to the formalisation of population genetic theory in its simple Mendelian system form.

Characterizing Mendelian genetics using a state space approach is more complicated. A theory on this view consists of the specification of a collection of

1990].

[23]The Gauss-Markov conditions provide a high level of confidence that the estimate of the parameters is reliable. Hence, it allows a determination of when least squares is a good method. The Gauss-Markov conditions are:

$$E(\varepsilon_i) = 0, \text{ for all } i$$
$$\text{var}(\varepsilon_i) = E(\varepsilon i - E(\varepsilon_i))^2 = E(\varepsilon_i^2) = \sigma^2$$
$$E(\varepsilon_i \varepsilon_j) = 0 \text{ for all } i \neq j$$

where ε_is and ε_js are errors, E is expectation and σ is variance.

mathematical entities (numbers, vectors, functions) used to represent states of the system within a state space (a topological structure), the behaviour of which is specified by three functions. These functions are commonly called laws but are not the same as laws in the received view. In the received view, laws are statements describing the behaviour of entities in the world. In the semantic view, laws are mathematical descriptions of the behaviour of mathematical systems of mathematical entities. The three laws are: laws of coexistence (which specify the physically possible set of states of the system); Laws of succession (which specify the possible histories of the system); and laws of interaction (which specify the behaviour of the system under conditions of inputs from interaction with other systems). A theory also requires the specification of a set of measurable magnitudes (represented by a function defined on the state space). Statements which formulate propositions to the effect that a particular magnitude has a particular value at a particular time are elementary statements. A satisfaction function determines the sets of states which satisfy the assignment of a value to a physical magnitude.

For population genetic theory, the state space will be a Cartesian n-space where 'n' is a function of the number of possible pairs of alleles in the population. A law of coexistence to the effect that only alleles at the same locus can form pairs will select the class of physically possible pairs of alleles. States of the system (genotype frequencies of populations) are n-tuples of real numbers from zero to one and are represented in the state space as points. These are the measurable magnitudes. An example of a satisfaction function for the elementary statement "genotype Aa occurs with a frequency of 0.5" would specify the set of states in the state space that satisfy the statement. In this case the set of states would be a Cartesian $(n - I)$-space, which is a subset of the state space. For population genetic theory, a central law of succession is the Hardy–Weinberg law.

State spaces (phase spaces) provide a powerful tool for characterising the dynamics of a system topologically. A simple and clear example of this power is given by Kellert [1993]; Kellert, explicating the theoretical account of turbulence given by Lev Landau [1944], provides a state space representation of turbulence at increasing levels of turbulence in a river.

By way of concluding this exposition of the semantic account, let me summarise its central features:

1. Scientific theories are not linguistic but rather mathematical entities

2. Theories are not appropriately formalised in first-order predicate logic with identity.

3. The specification of the model(s) whose intended scope is an empirical system is a complete formalisation of the theory.

4. The process for confirming a theory is complex and extra-theoretical involving a justification of an assertion that the model is isomorphic with the empirical system that is its intended scope. This requires recourse to numerous other theories (mathematical and empirical).

5. Explanation consists in demonstrating that the dynamics of the model correspond to dynamics of the empirical system in need of explanation.

6. Prediction consists in exploring the dynamics of the model and asserting a future state of the empirical system.

In order to dispel a potential misunderstanding, I think it important to emphasize one crucial similarity between the syntactic and semantic conceptions, namely, they are both conceptions of the formal structure of theories. In addition, both conceive of theories as deductive systems and both require fundamental principles (axioms) in the formulation of the dynamics of a system. Hence, there is no refuge to be found in the semantic conception for those philosophers who dispute the appropriateness, and usefulness of formalisation in one or all branches of science. Those of us who espouse the semantic conception are, like those who espouse the syntactic conception, committed to the value of formalisation in science and philosophy of science.

1.3 The Galilean Conception of Scientific Theories

Patrick Suppes encapsulates his view of the formalisation of scientific theories as follows:

> The sense of formalization I shall use in the subsequent discussion is just that of a standard set-theoretical formulation. I do not want to mean by formalization the stricter conception of a first-order theory that assumes only elementary logic. Such stricter formalization is appropriate for the intensive study of many elementary domains of mathematics, but in almost all areas of science a rich mathematical apparatus is needed. We can properly appeal to that apparatus within a set-theoretical framework [Suppes, 1968, p.653].

As the state-space version of the semantic view demonstrates, we can properly appeal to that apparatus with a topological framework as well. Indeed, confining the "rich mathematical apparatus" to set theory or topology is unnecessarily restrictive and does not accord with scientific theorising in many domains. Opening up the domain of mathematics that can be drawn upon is a natural extension of the semantic view.

As indicated above, by dubbing an extension of the semantic conception, the Galilean conception, I am explicitly connecting it to Galileo's famous claim (1623):

> Philosophy is written in this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.

On this conception of scientific theories, the empirical world in all its complexity is formalised in the language of mathematics with all its richness.[24] Hence, the formalisation of a theory rests on finding the most appropriate domain of mathematics and mathematical techniques. As van Fraassen[25] has pointed out, the slogan that motivated Suppes' conception of scientific theories was that philosophy of science should use mathematics not meta-mathematics.[26]

The broadest and simplest definition of a mathematical model[27] construes it as a specification, using a domain of mathematics, of the kinds of entities involved in a system and the dynamics of their behaviour. Things are more complicated than this definition captures, however, because a mathematical model *of empirical phenomena* will always involve an idealisation of the phenomenal world.

There are at least four major benefits of conceiving of theories as mathematical models.[28] First, a mathematical model introduces precision which, in part, results from a removal of the ambiguity that plagues ordinary language — however, part of the cost of precision is idealization and a high level of metrication. Second, mathematical models provide a powerful machinery for exploring the dynamics of the phenomena. From now on, I will refer to the collection of properties of a class of phenomena as a phenomenal system, and to the collection of properties of a mathematical model as a system (a designation which is short for model system). Hence, the expression, "dynamics of the system" refers to the dynamics of the model which is purported to represent the dynamics of a phenomenal system.

[24] If one accepts logicism (the view that all of mathematics can be reduced to deductive logic), then, in principle, any mathematical model can be reduced to a formalisation in logic. The most well known advocate of logicism was Bertrand Russell. Logicism is still the subject of debate. For the most part, the major challenge to logicism was Gödel's discovery that the derivation of all mathematical truths requires a logic that it is impossible to formalise. Although an important issue in the foundations of mathematics, whether all of mathematics can be reduced to logic is too esoteric an issue in the context of the formalisation of scientific theories. Even if logicism is correct, the complexity of such a reduction will be great but, more importantly, it will provide no added value to actual scientific theorising.

[25] van Fraassen [1980, p. 65].

[26] Meta-mathematics, following Tarski [1935; 1936], encompasses the semantic and syntactical examination of formal languages and systems. It is the logical examination of rules specifying the permissible combinations of symbols and numbers in mathematics and their uses, as well as the logical examination of mathematical methods and principles. First-order predicate logic with identity figures prominently in the study of the foundation of mathematics and has become the general study of the logical structure of axiomatic systems. Hence, appeal to first-order predicate logic in the formalisation of scientific theories is using meta-mathematics, not mathematics.

[27] It is important to be precise about this because the term "model" in science (and other contexts) is open to numerous interpretations. A mathematical model, in the context of the formalisation of a scientific theory, is the formulation of the dynamics of a system using the formal calculus of a domain of mathematics that embodies a specific interpretation of the calculus. The formulation of the dynamics in such a language consists of specifying the formulae that encode all the relevant dynamical properties. Sometimes this may be a single equation or a small set of equations such as the three equations used by Lorenz to formalise the dynamics of turbulence [Lorenz, 1963].

[28] The first benefit is also realized within the syntactic conception. The remaining three are also realized within the semantic conception but less fully. This is not surprising since the Galilean conception is merely an extension of the semantic conception.

Third, links among different phenomenal domains become apparent. This is a more abstract benefit than the first two. Consider, however, the application, alluded to above, of game theory to economics, ecosystems, and international politics. Providing a model in terms of game theory for each of these domains established the similarity of the dynamics in each area and a cross-fertilization of ideas became possible. Fourth, the coherence of the knowledge of phenomena is achieved; a coherence which, when the knowledge is expressed in a large and complex number of ordinary language sentences, is less obvious and certainly less demonstrable.

1.4 Formalisations and Theories

To this point, the relationship among the terms mathematical model, formalisation and theory has not been explicitly described. It will be useful to provide such a description, as well to be clear on what constitutes a scientific theory, prior to looking at the nature of evolutionary theory and some examples of formalisations of elements of it.

Although a theory can be given a formalisation, not all formalisations are formalisations of a scientific theory (consider Hilbert's axiomatisation of Euclidean geometry). Also, not all formalisations employ mathematical models; in the logical empiricist view of theories, theories were not formalised as mathematical models in the common sense of a mathematical model. In addition, not all mathematical models are formalisations of scientific theories. What, then, is the relationship among these concepts? What I have argued, is that the most adequate understanding of scientific theories holds that they are formalised by constructing mathematical models of dynamical systems, where the mathematical models are claimed to be isomorphic (homeomorphic) to the empirical system of phenomena within their intended scope.

If, however, not all mathematical models are, or are intended to be, a formalisation of a scientific theory, what distinguishes a formalisation of a scientific theory from other mathematical models? The answer to this question appeals to function. A scientific theory attempts to:

- integrate knowledge,
- facilitate a conceptual exploration of a dynamical system to discover unknown (an perhaps empirically unknowable) properties (such as the Lorenz attractor in system of turbulence),
- render phenomena in a system explainable
- allow predictions
- guide future research (e.g., hypothesis formation, data interpretation)

The central feature of a scientific theory is its intended application to an empirical system. *A key element of scientific theorising is, therefore, the complex*

relationship between the model and the empirical system within its intended scope. It is that relationship and not the mathematical model which is the essence of a ***scientific*** *theory.* It is also a reason why one mathematical model can participate in scientific theorising in a number of different domains (e.g., game theory which has been applied to economic behaviour[29], ecology[30] and social structure[31]).

The value and power of mathematical models resides in three characteristics: abstraction, systematisation and deduction. Constructing a mathematical model involves abstracting the essential elements of a system's ontology and dynamics and constructing a conceptual framework. Once a model has been constructed, the deductive machinery of the particular domain of mathematics is available to allow an exploration of the properties of the system. Sunny Auyang has expressed these features concisely:

> More important than calculation is the power of mathematics to abstract and articulate precisely ideas with wide implications, to construct complicated conceptual structures, and to analyse exactly the interconnection of their elements.[32]

2 THE COMPLEX STRUCTURE OF EVOLUTIONARY THEORY

I have argued elsewhere that evolutionary theory is a composite of a number of interacting sub-theories [Thompson, 1983; 1986; 1989]. Interestingly, this idea was first suggested by Morton Beckner, 'A "theory" may be a family of models' [1959, 53–54]. This insight was lost for the next two decades as philosophers interested in biology focused on the logical empiricist conception of theories — some attempting to show that biological theories did not fit the schema and others arguing that it did. With the advent of the semantic conception of theories and the exploration of its applicability to biology, a return to Beckner's insight became possible.

In my earlier works, I suggested that the interacting theories were population genetics, selection theory, and ecology. It now clear that this was too simplistic. A complete evolutionary theory also requires a theory of developmental, and of cell and molecular biology. As is apparent in section III below, of the various domains comprising evolutionary theory, some have a rich mathematical formalisation (e.g., population genetics, ecology), others have a more embryonic formalisation (e.g., developmental biology[33]).

A somewhat crude characterisation of the interrelationship of the component domains that comprise evolutionary theory can be given by employing the concept, "partially determines" where "A partially determines B if the dynamics of A affect the dynamics of B". Beginning at the macro-level, the following relations

[29] See any major economics text.
[30] See: [Bulmer, 1994, ch. 8; Maynard Smith, 1982; 1983].
[31] See, [Skyrms, 1996; 2004].
[32] [Auyang, 1998].
[33] A measure of the degree of formalisation is the ability to provide a state space representation of the dynamics and trajectories

of "partially determines" obtain. Ecological dynamics (including the phenotypic properties of individuals of a population at generation F_0) partially determines selection dynamics on a population at F_0. Selection dynamics on F_0 partially determines the population genetic structure of the breeding population and hence the population dynamics resulting in the next generation (F_1). Population genetic dynamics partially determines the molecular genetic structure of the population at F_1. The molecular genetic dynamics of F_1 partially determines the developmental dynamics. The developmental dynamics partially determines the phenotype of F_1.

The population genetics component of evolutionary theory was the earliest to be formalised as a result of the work of Ronald A. Fisher, John B.S. Haldane and Sewall Wright. Fisher published the most comprehensive initial account of population genetics, *The Genetical Theory of Natural Selection* [1930]. A superb contemporary account is found in, *Principals of Population Genetics* (3^{rd} ed.) by Daniel Hartl and Andrew Clark [1997].[34] The richness of Sewall Wright's contributions were not fully appreciated until 1960's.[35]

As indicated earlier, Mary Williams' [1970] axiomatisation of selection theory fits naturally into this family of mathematical models conception of evolutionary theory. She has provided an excellent set-theoretical axiomatisation of a crucial component. Alone, contrary to Rosenberg, it is not a formalisation of evolutionary theory. In section 3.2, I provide a sketch of her formalisation of selection theory and indicate why it constitutes a Galilean conception formalisation, namely, that it formalises selection theory as a set-theoretical model. The third component that I sketch in section 3 is population ecology.

Returning to the integrative nature of evolutionary theory, one of the marks of a successful theory is its ability to integrate previously disparate bodies of knowledge. The power of Darwin's theory of evolution was its ability to integrate observations of biogeographic distributions, the geological/paleontological observations, anatomical observations (homologies and vestiges), observations of the remarkable adaptedness of most organisms to their environment, etc. This consilience of inductions, in Whewell's terminology, is a manifestation of the power of the theory. Biological evolution, in all its grandeur, conforms to this *desideratum* wonderfully. As will be seen below, the variety of mathematical models employed in evolutionary theory, though integrated and made plausible by the conceptual framework of evolution, draw on different domains of mathematics. And, recently, with the increase in computing power available on the desktop and powerful pro-

[34]Hartl has also produced an excellent primer: *A Primer of Population Genetics* (3^{rd} ed.) [2000].

[35]For an excellent biography of Sewell Wright and an exposition of his contributions to genetics, see Provine [1986a]. William Provine has also edited a collection of Wright's papers [Provine, 1986b]. In the Preface he remarks that "Wright's papers, especially those published before about 1950, ... were little understood at the time of publication, even if widely read." Provine's explanation for this is: "Evolutionary biologists in general had very little training in mathematics or specifically in statistics, or in quantitative reasoning generally. Moreover, Wright was rather insensitive to the inability of his audience to follow his quantitative reasoning. Even those with some mathematical training had much difficulty following Wright's idiosyncratic method of path coefficients. ... Now, however, the situation has changed dramatically."

gramming languages like *Mathematica* and *Maple*, the dynamics of systems have be formalised algorithmically and the algorithms "run" to produce a computer simulation of the behaviour.[36]

3 THE FORMALISATION OF SOME COMPONENT THEORIES WHICH COMPRISE CONTEMPORARY EVOLUTIONARY THEORY

3.1 Formalisation in Selection Theory

I have in various places above indicated that Mary William's formalisation of selection theory is not a syntactic conception formalisation. It is an axiomatisation and draws on the deductive machinery of mathematics:

> It shows this by stating them [the seven fundamental principles she identifies as implicit in Darwin's *Origin of Species*] in a sufficiently rigorous form to allow the deductive techniques of ***mathematics*** to be used to derive from them other principles of the theory (Williams p. 344 italics added for emphasis)

Deduction and axiomatisation, however, are features of the syntactic, semantic and Galilean conceptions. This is quite explicit in Patrick Suppes' set-theoretical predicate version of the semantic view but is also true of topological versions. What is specific to the syntactic conception is the requirement that the axiomatisation be in first-order predicate logic with identity and that semantics are provided by correspondence rules. Even if one weakens the conception to permit the semantics to be given using models, it is the requirement that the axiomatisation be in first-order predicate logic that separates it from the semantic and Galilean conceptions.

Mary William's axiomatisation is not in first-order predicate logic. It is in set theory, and the deductive apparatus employed is that of set theory. To anyone who studies her formulation of the fundamental principles and their deductive consequences, it will be clear that she has provided a set theoretical axiomatisation. But even a cursory examination of the paper makes this clear. Consider the following:

> *Informal Theorem B 3*
> Every theorem that mathematicians have proved about strict partial ordering is automatically a true theorem about the ancestor relation. Therefore biologist's knowledge about the ancestor relation can be enriched by mathematician's knowledge about strict partial orderings. (p. 348)

The definition of "strict partial ordering" that she employs is from set theory and the specific definition she uses is from Patrick Suppes' *Axiomatic Set Theory* [1960]:

[36] See, for example, [Gaylord and Nishidate, 1996].

> The primary concept of set theory, that of a set, has already been used extensively without any explicit definition. This was possible because the meaning of set within set theory is very close to its meaning in its everyday usage (and it was necessary because *set* is a primitive term of set theory and hence cannot be defined explicitly). However, in order to go further in this discussion of subclans, I will need to use a few notions from set theory which are not self explanatory; therefore, this section will be used to describe or define them. (p.353)

She then describes or defines: null set, subset, intersection, union and relative complement.

Consequently, her formalisation is more appropriately regarded as a semantic or Galilean conception formalisation. Once it is seen as a semantic conception of theories formalisation, it can be regarded properly as a powerful example of the benefits of formalisation and as one of the family of models that make up evolutionary theory. Although Williams uses the term "evolution" in the title of her paper, she is quite explicit that she is formalising Darwin's theory of evolution as set out in the *Origin of Species*.

> Here, and throughout the paper, I use "Darwin's theory of evolution" to denote *the part of the present theory* for which Darwin provided all of the crucial explanatory principles; briefly this is the theory of anagenesis, or descent with adaptive modification. (p. 344 italics added for emphasis)

As the italicised phase makes clear, Williams acknowledges that this is only part of the present theory of evolution.

What follows is a summary; a complete presentation of her formalisation would require duplicating most of her paper. Those interested in her impressive formalisation would be better served by going to the paper itself. A flavour of the elements of the formalisation, however, can be given easily since Williams provides both an informal and a formal (mathematical) statement of each of the axioms and theorems which she deduces from them. The formalisation and its value reside in the mathematical formulation but the informal statement captures the essence of her approach. The informal statement of the seven axioms (fundamental principles) is as follows (although the sequence corresponds to that in the paper, the actual numbering does not):

1. No biological entity is a parent of itself.
2. If b_1 is an ancestor of b_2, then b_2 is not an ancestor of b_1.
3. Every Darwinian subclan is a subclan of a clan in some biocosm;
    - the clan of a set S is the set of all its descendants;
    - a subclan is either a whole clan or a clan with several branches removed;

- a biocosm is a set of subclans in which change through time is directed by the laws of natural selection.

4. There is an upper limit to the number of organisms in any generation of a Darwinian subclan.
5. For each organism there is a positive real number which describes its fitness in its particular environment.
6. Consider a subcland D_1 of D. If D_1 is superior in fitness to the rest of D for sufficiently many generations (where how many is "sufficiently many" is determined by how superior D_1 is and how large D_1 is), then the proportion of D_1 in D will increase;

    - a subcland is a subclan which is contained within a subclan (i.e., a sub-subclan)

7. In every generation m of a Darwinian subclan D which is not on the verge of extinction, there is a subcland D_1 such that: D_1 is superior to the rest of D for long enough to ensure that D_1 will increase relative to D; and as long as D_1 is not fixed in D it retains sufficient superiority to ensure further increases relative to D.

In their mathematical formulation, these seven axioms allow numerous theorems to be deduced. For example (again expressed informally):

- No biological entity is its own ancestor.
- If $C_1(m)$ is a subset of $C_2(m)$, then for all subsequent generations k, $C_1(k)$ must be a subset of $C_2(k)$.
- The fitness of an organism relative to a subclan to which it belongs is always positive.
- For every Darwinian subclan D which does not die out, there is an infinite sequence of subclands, D_1, D_2, ..., such that each subcland is contained in it predecessor and each is fitter than its predecessor for long enough for natural selection to ensure that it becomes fixed in D.

In all Williams derives 20 theorems from the axioms. The utility of formalisation is demonstrated by her ability to provide a rigorous deductive proof for each of the theorems once a mathematical (set theoretical) formulation of each of the axioms is given. As she correctly points out, many more theorems could be proved but the 20 she provides more than justify the claim she and, in similar words, many others, including me, have made:

> The value of such a deductive system (or axiomatisation) is that it provides a technique for discovering new phenomena which are too remote from observation or too unintuitive to ever be discovered with the unaided intuition... . (p.343)

3.2 Formalisation in Population Genetics

The phenomenon of heredity, although widely accepted since at least the Greco-Roman period, is extremely complex and an adequate theory proved allusive for several thousand years. Indeed, features of heredity seemed almost magical. Breeders from antiquity had a sophisticated understanding of the effects of selective breeding but even the most accomplished breeders found many aspects of heredity to be capricious. Even Darwin in the middle 19^{th} century knew well the techniques of selective breeding (artificial selection) but did not have available a satisfactory theory of heredity when he published the *Origin of Species* [1859]. Although, he realized that his theory of evolution depended on heredity, he was unable to provide an account of it. Instead, he relied on the widely known effects of artificial selection and by analogy postulated the effects of natural selection in which the culling of breeders was replaced by forces of nature.

The first major advance came from the simple experiments and mathematical description of the dynamics of heredity by Gregor Mendel [1865]. Although Mendel's work went largely unnoticed until the beginning of the 20^{th} century, its great strength lay in its mathematical description — elementary though that description was. Mendel performed a number of experiments which provided important data but it was his elementary mathematical description of the underlying dynamics that has had a lasting impact on genetics. His dynamics were uncomplicated. He postulated that a phenotypic characteristic (characteristic of organisms) is the result of the combination of two "factors" in the hereditary material of the organism. Different characteristics are caused by different combinations. Focusing on one characteristic at a time made the problem of heredity tractable. Factors could be dominant or recessive. If two dominant factors combined, the organism would manifest the characteristic controlled by that factor. If a dominant and a recessive factor combined, the organism would manifest the characteristic of the dominant factor (that is the sense in which it is dominant). If two recessive factors combine, the organism will manifest the characteristic of the recessive factor.

Mendel postulated two principles (often now referred to as Mendel's *laws*): a principle of segregation and a principal of independent assortment. The principle of segregation states that the factors in a combination will segregate (separate) in the production of gametes. That is gametes will contain only one factor from a combination. The principle of independent assortment states that the factors do not blend but remain distinct entities and there is no influence of one factor over the other in segregation. The central principle is the law of segregation. The law of independent assortment can be folded into the law of segregation as part of the definition of segregation. When gametes come together in a fertilised ovum (a zygote), a new combination is made.

Assume A is a dominant factor and a is a recessive factor. Three combinations are possible AA, Aa and aa. Mendel's experimental work involved breeding AA plants and aa plants. He then crossed the plants which produced only Aa plants. He then bred the Aa plants. What resulted was $.25AA$, $.5Aa$ and $.25aa$. His

dynamics explains this result. Since the factors A and a do not blend and they segregate in the gametes and combine again in the zygote, the results are fully explained. Crossing the AA plants with aa plants will yield only Aa plants:

		Sperm	
		A	A
Ova	a	Aa	Aa
	a	Aa	Aa

Breeding only Aa plants will yield the .25:.5:.25 ratios:

		Sperm	
		A	a
Ova	A	AA	Aa
	a	Aa	aa

Two of four cells yield Aa that is .5 of the possible combinations. Each of AA and aa occupy only one cell in four, that is, .25 of the possible combinations. In contemporary population genetics, Mendel's factors are called alleles. The location on the chromosome where a pair of alleles is located is called a locus. Sometime the term gene is used as a synonym for allele but this usage is far too loose. Subsequently, I will explore the confusion, complexity and controversy over the definition of "gene." Mendel's dynamics assumed diallelic loci: two alleles per locus. His dynamics are easily extended to cases where each locus has many alleles any two of which could occupy the locus.

The basic features of Mendel's dynamics were modified and extended early in the 20^{th} century. G. Udny Yule [1902] was among the first to explore the implications of Mendel's system for populations. In a verbal exchange between Yule and R.C. Punnett in 1908, Yule asserted that a novel dominant allele arising among a 100% recessive alleles would inexorably increase in frequency until it reach 50%. Punnett believing Yule to be wrong but unable to provide a proof, took the problem to G. H. Hardy. Hardy, a mathematician, quickly produced a proof by using variables where Yule had used specific allelic frequencies. In effect, he developed a simple *mathematical model*. He published his results in 1908. What emerged from the proof was a principle that became central to population genetics, namely, after the first generation, allelic frequencies would remain the same for all subsequent generations; an equilibrium would be reached after just one generation. Also in 1908, Wilhelm Weinberg published similar results and articulated the same principle (the original paper is in German, and English translation is in Boyer [1963]). Hence, the principle is known as the Hardy-Weinberg principle or the Hardy-Weinberg equilibrium.[37] In parallel with these mathematical advances was

[37]William Castle [1903] also generated the equilibrium principle using numerical analysis but

a confirmation of the phenomenon of segregation and recombination in the new field of cytology.

Building on this early work, a sophisticated mathematical model of the complex dynamics of heredity emerged during the 1920s and 1930s, principally through the work of John Haldane [1924; 1931; 1932], Ronald Fisher [1930] and Sewall Wright [1931; 1932]. What has become modern population genetics began during this period. From that period, the dynamics of heredity in populations has been studied from within a mathematical framework.[38]

As previously indicated, one of the fundamental principles of the theory of population genetics, in the form of a mathematical model, is the Hardy-Weinberg Equilibrium. Like Newton's First Law, this principle of equilibrium states that after the first generation if nothing changes then allelic (gene) frequencies will remain constant. The presence of a principle(s) of equilibrium in the dynamics of a system is of fundamental importance. It defines the conditions under which nothing will change. All changes, therefore, require the identification of cause(s) of the change. Newton's dynamics of motion include an equilibrium principle that states that in absence of unbalanced forces an object will continue in uniform motion or at rest. Hence acceleration, deceleration, change of direction all require the presence of an unbalanced force. In population genetics, in the absence of some perturbing factor, allelic frequencies at a locus will not change. Factors such as selection, mutation, meiotic drive, and migration are all perturbing factors. Like many complex systems, population genetics also has a stochastic perturbing force, commonly call random genetic drift.

In what follows, the central features of the mathematical model of contemporary population genetic theory are set out. Quite naturally, the exposition begins with the Hardy-Weinberg Equilibrium. It is useful to begin with the exploration of a one locus, two-allele system. In anticipation, however, of multi allelic loci, we switch from A and a to 'A_1' and 'A_2'. Hence, according to the Hardy-Weinberg Equilibrium, if there are two different alleles 'A_1' and 'A_2' at a locus and the ratio in generation 1 is $A_1{:}A_2 = p{:}q$, and if there are no perturbing factors, then in generation 2, and in all subsequent generations, the alleles will be distributed:

$$(p^2)A_1A_1 : (2pq)A_1A_2 : (q^2)A_2A_2.$$

The ratio of $p : q$ is normalised by requiring that $p + q = 1$. Hence, $q = 1 - p$ and $1 - p$ can be substituted for q at all occurrences. The proof of this equilibrium is remarkably simply.

The boxes contain zygote frequencies. In the upper left box, the frequency of the zygote arising from the combination of an A_1 sperm and A_1 egg is $p \times p$, or p^2, since the initial frequency of A_1 is p. In the upper right box, the frequency of the zygote arising from the combination of an A_2 sperm and A_1 egg is $p \times q$, or pq, since the initial frequency of an A_2 is q and the initial frequency of A_1 is p.

his name is not often used in connection with the principle, probably because his results were not in an abstract generalized form.

[38]For an excellent history of the development of population genetics, see Provine [1971].

		Sperm	
		$\mathrm{fr}(A_1) = p$	$\mathrm{fr}(A_2) = p$
Ova	$\mathrm{fr}(A_1) = p$	$\mathrm{fr}(A_1A_1) = p^2$	$\mathrm{fr}(A_2A_1) = pq$
	$\mathrm{fr}(A_2) = P$	$\mathrm{fr}(A_1A_2) = pq$	$\mathrm{fr}(A_2A_2) = q^2$

The lower left box also yields a pq frequency for an A_1A_2. Since the order doesn't matter, A_2A_1 is the same as A_1A_2 and hence the sum of frequencies is $2pq$.

This proves that a population with $A_1 : A_2 = p{:}q$ in an initial generation will in the next generation have a frequency distribution: $(p^2)A_1A_1 : (2pq)A_1A_2 : (q^2)A_2A_2$. The second step is to prove that this distribution is an equilibrium in the absence of perturbing factors. Given the frequency distribution $(p^2)A_1A_1 : (2pq)A_1A_2 : (q^2)A_2A_2, p^2$ of the alleles will be A_1 and half of the A_1A_2 combination will be A_1, that is pq. Hence, there will be $p^2 + (pq)A_1$ in this subsequent generation. Since $q = (1 - p)$, we can substitute $(1 - p)$ for q, yielding $p^2 + (p(1-p)) = p^2 + (p - p^2) = p$. Since the frequency of A_1 in this generation is the same as in the initial generation (i.e., p), the same frequency distribution will occur in the following generation (i.e., $(p^2)A_1A_1 : (2pq)A_1A_2 : (q^2)A_2A_2$).

Consequently, if there are no perturbing factors, the frequency of alleles after the first generation will remain constant. But, of course, there are always perturbing factors. One central one for Darwinian evolution is selection. Selection can be added to the dynamics by introducing a coefficient of selection. For each genotype (combination of alleles at a locus[39]) a fitness value can be assigned. Abstractly, A_1A_1 has a fitness of W_{11}, A_1A_2 has a fitness of W_{12}, and A_2A_2 has a fitness of W_{22}. Hence, the ratios after selection will be:

$$W_{11}(p^2)A_1A_1 : W_{12}(2pq)A_1A_2 : W_{22}(q^2)A_2A_2.$$

To calculate the ratio $p : q$ after selection this ratio has to be normalised to make $p + q = 1$. To do this, the average fitness, ϖ, is calculated. The average fitness is the sum of the individual fitnesses.

$$\varpi = w_{11}(p^2) + w_{12}(2pq) + w_{22}(q^2).$$

Then each factor in the ratio is divided by ϖ, to yield:

$$((w_{11}(p^2))/\varpi)A_1A_1 : ((w_{12}(2pq)/\varpi)A_1A_2 : ((w_{22}(q^2)/\varpi)A_2A_2.$$

[39]The term "genotype" is used in a number of different ways by different writers and in different contexts. It can, as in this context, mean a pair of alleles at a locus. It is sometimes used to designate an organism's entire collection of loci and the alleles associated with them (see, for example, [Hartl, 2000, p.2]). It is also used to designate a subset of loci that together determine some characteristic of the organism. This variety of uses invites confusion but has now become embedded in the language of evolutionary biologists, geneticists and ecologists. Hence, to avoid problems that can arise from the equivocation on the term, clarity about the use in particular contexts is essential.

Other factors such as meiotic drive can be added either as additional parameters in the Hardy-Weinberg equilibrium or as separate ratios or equations.

Against this background, a precise application of a X^2-test of goodness of fit can be provided. The following example[40] illustrates the determination the goodness of fit between observed data and the expected data based on the Hardy-Weinberg equilibrium. The human chemokine receptor[41] gene CC-CKR-5codes for a major macrophage co-receptor for the human immunodeficiency virus HIV-1. CC-CKR-5 is part of the receptor structure that allows the entry of HIV-1 into macrophages and T-cells. In rare individuals, a 32-base-pair indel[42] results in a non-functional variant of CC-CKR-5. This variant of CC-CKR-5 has a 32-base-pair deletion from the coding region. This results in a frame shift and truncation of the translated protein. The indel results when an individual is homozygous for the allele $\Delta 32$[43]. These individuals are strongly resistant to HIV-1; the variant CC-CKR-5 co-receptor blocks the entry of the virus into macrophages and T-cells.

In a sample of Parisians studied for non-deletion and deletion (+ and $\Delta 32$ respectively), Lucotte and Mercier (1998) found the following genotypes:

$$++ : 224 \quad + \Delta 32 : 64 \quad \Delta 32 \Delta 32 : 6$$

Dividing by the populations sample size yields the genotype frequencies:

$$++ : 224/294 = 0.762 \quad + \Delta 32 : 64/294 = 0.218 \quad \Delta 32 \Delta 32 : 6/994 = 0.20$$

Multiplying the number of homozygotes for an allele by 2 and adding the number of heterozygotes yields the number of that allele in the sample. Dividing that by

[40] The data is taken from Hartl [2000].

[41] Immune responses depend on the activities of white blood cells (leukocytes). Some immune responses are innate (i.e., one is born with them). These depend, to a large extent, on the action of granulocytes of which macrophages are an instance. Some immune responses are adaptive (i.e., one develops the immunity in response to a pathogen). These depend on lymphocytes of which T lymphocytes (or T-cells) are an instance. Macrophages and T-cells have on their surfaces highly diverse receptors. These receptors can be stimulated to provoke an action or can be used to bind to a pathogen. Sometimes binding to a pathogen is sufficient to block its entry into a cell and, hence, neutralize it. Sometimes binding enables a phagocyte cell to ingest the pathogen. And there are a large number of other mechanisms. A cytokine is a protein made by cells, that activates a behaviour in another cell. A chemokine is a small cytokine, prinicipally involved in the migration and activation of phagocytes and lymphocytes. The chemokine receptor on a macrophage is the receptor through which chemokines stimulate the activity of the macrophage. It is also the receptor that provides an entry point into the macrophage of HIV-1.

[42] DNA in an intertwined helical ladder, the rungs of which consist of two connected nucleotides. That is each rung is comprised of two joined nucleotides. There are four nucleotides in DNA: adenine (A), guanine (G), thymine (T) and cytosine (C). the only possible combinations are C and G, and A and T. So a segment of five rungs could be A-T, T-A, G-C, A-T, C-G. DNA polymorphisms are ubiquitous in nature. That is, at a defined site in the DNA of an organism, many different sequences of nucleotide pairs are present in different individuals (nucleotide polymorphism). An indel is an insertion or deletion polymorphism – a sequence of nucleotides are added or deleted in some individuals creating a polymorphism at that site. A 32-base-pair indel is an addition/deletion of 32 rungs of the DNA ladder.

[43] The reference to $\Delta 32$ is a switch in reference points. CC-CKR-5 is a nucleotide sequence at the molecular-genetic-level/DNA. $\Delta 32$ is an allele at the population-genetic-level/chromosomal level. This allelic composition determines the DNA composition of the individual.

the sample size times 2 (there are twice as many alleles as individuals) yields the allelic frequency of this sample. Hence:

> The frequency of the + allele = 0.871
> The frequency of the $\Delta 32$ allele = 0.129

What genotype numbers does the hardy-Weinberg equilibrium yield given these allelic frequencies?

$$\begin{array}{ll} & (p^2) + + : (2pq) \quad + \Delta 32 : (q^2) \quad \Delta 32 \Delta 32 \\ \text{Yields} & (0.871^2) + + : (2(0.871X0.129)) + \Delta 32 : (0.129^2) \Delta 32 \Delta 32 \\ = & 0.758641 + + : 0.224718 + \Delta 32 : 0.016641 \Delta 32 \Delta 32 \end{array}$$

Hence, in a population of 294 individuals, the Hardy-Weinberg equilibrium yields:

$$++ : 222.9 \quad + \Delta 32 : 66.2 \quad \Delta 32 \Delta 32 : 4.9$$

As we would expect these add up to 294. A comparison of the values expected based on the Hardy-Weinberg equilibrium and those observed yields:

$$\begin{array}{ll} \text{H} - \text{D expected :} & ++ : 222.9 \quad + \Delta 32 : 66.2 \quad \Delta 32 \Delta 32 : 4.9 \\ \text{Observed :} & ++ : 224 \quad + \Delta 32 : 64 \quad \Delta 32 \Delta 32 : 6 \end{array}$$

Now we can ask, how good is the fit between the H-D expected values based on the specified allelic frequencies and the observed values?

The X^2-test is:

$$X^2 = \sum \text{ (observed quantity – expected quantity)}^2/\text{(expected quantity)}$$

There are three genotypes, hence:

$$\begin{array}{rcl} X^2 & = & ((224 - 222.9)^2/222.9) + ((64 - 66.2)^2/66.2) + ((6 - 4.9)^2/4.9) \\ & = & (1.21/222.9) + (4.84/66.2) + (1.21/4.9) \\ & = & 0.00543 + 0.0731 + 0.2469 \\ & = & 0.3254 \end{array}$$

To use this result to assess goodness of fit, it is necessary to determine the degrees of freedom for the test.

> Degrees of Freedom (df) = (classes of data - 1) – the number of parameters estimated.

Since there are three genotypes, the classes of data is 3. Since $p + q = 1$(hence, q is a function of p; they are not independent parameters), there is only 1 parameter being estimated. Hence, the degrees of freedom for this test is:

$$(3 - 1) - 1 = 1$$

Using the X^2 result and 1 degree of freedom allows a probability value to be determined.

In this case, the relevant probability is 0.63. This is the probability that chance alone could have produced the discrepancy between the H-D expected values and the observed values. Since we are measuring the probability that chance alone could have produced *the discrepancy* (not to be confused with *the similarity* between the two[44]), the higher the probability, the more robust one's confidence that there are *no factors other than chance* causing the discrepancy and, hence, that there is a good fit between the values expected based on the model and the observed values[45]; any discrepancy is a function of chance alone.

The elementary framework sketched above has been expanded to include the Wright-Fisher model of Random Drift, mutations, inbreeding and other causes of non-random breeding, migration speciation, multiple alleles at a locus, multi-loci systems, phenotypic plasticity, etc. One important expansion relates to interdemic selection.

The account so far describes intrademic selection. That is, selection of individuals within an interbreeding population — a deme. However, the mathematical model also permits the exploration of interdemic selection (selection between genetically isolated populations) using adaptive landscapes. One outcome of such explorations is a sophisticated account of why and how populations reach sub-maximal, sub-optimal peaks of fitness. Richard Lewontin, building on concepts set out by Sewell Wright, provided the first mathematical description of this phenomenon.

Consider a population genetic system with two loci and two alleles (here for simplicity I revert to upper and lower case letter for alleles and for dominance and recessiveness). The possible combinations of alleles is:

	AB	Ab	aB	ab
AB	$AABB$	$AABb$	$AaBB$	$AaBb$
Ab	$AABb$	$AAbb$	$AaBb$	$Aabb$
aB	$AaBB$	$AaBb$	$aaBB$	$aaBb$
ab	$AabBb$	$Aabb$	$aaBb$	$aabb$

[44]If one were to determine the probability that the *similarity* was due to chance alone, the lower the probability the more robust the isomorphism. That is, the lower the probability that chance alone has produced an agreement of the two domains, the higher the probability that the agreement is due to the intrinsic features of the domains (e.g., the ontology and causal structure).

[45]Determining at precisely what probability value one is justified in declaring discrepancies are important (significant) is controversial. A convention has emerged that a $P < .05$ cutoff yields statistical significance for the discrepancies. That is, if the probability that the discrepancy is due to chance is less than .5 it is reasonable to reject a claim that there is a fit between the model and the observed data — the discrepancy is significant and cannot be dismissed.

There are 9 different combinations (genotypes). For each genotype a fitness co-efficient W_i can be assigned. In addition, for each genotype a frequency can be assigned based on p_1 and q_1, p_2 and q_2 (for locus 1 and locus 2 respectively). Let that frequency be Z_i. The product of the frequency of a genotype and the fitness of that genotype is the contribution to the average fitness of the population ϖ made by that genotype. The sum of the contributions of all the genotypes represented in the population is the average fitness ϖ of the population. Hence, the average fitness ϖ for a population

$$= \sum Z_i W_i.$$

Consider the following calculation for a single population.

$$P_1 = .7$$
$$p_2 = .4$$

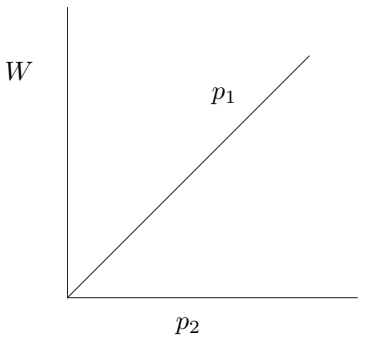
Since $p_1 + q_1 = 1$ and $p_2 + q_2 = 1$, the value of q can be determined from the value of p. Hence the value of p alone is sufficient to determine the genotype frequencies of the population.

In accordance with the Hardy-Weinberg equilibrium, the genotype frequencies can be calculated by multiplying the frequencies of the allelic combinations at each locus in the two loci pair. The resulting frequencies with assigned fitnesses, frequency-fitnesses, and the average fitness for the population is shown in the following table:

Genotype	Frequency Z	Fitness W	Frequencey × Fitness
$AABB$	0.784	0.85	0.06664
$AABb$	0.23522	.0.48	0.108192
$AAbb$	0.1764	0.54	095256
$AaBB$	0.0672	0.87	0.058484
$AaBb$	0.2016	0.65	0.13104
$Aabb$	0.1512	0.32	0.048384
$aaBB$	0.0144	0.61	0.008784
$aaBb$	0.0432	1.2	0.05184
$aabb$	0.0324	1.13	0.036612
		$w =$	0.605212

By plotting the average fitness ϖ of each possible population in a two loci system with the assigned fitness values W_i, an adaptive landscape for the system can be

generated. This adaptive landscape is a three dimensional phase space (a system with a larger number of loci will have a correspondingly larger dimensionality):



The plotted point is the average fitness of the population described above. A complete adaptive landscape is a surface with adaptive peaks and valleys. An actual population under selection may climb a slope to an adaptive peak that is sub maximal (i.e., the average fitness of the population is less than the highest average fitness in the system). The only way to move to another slope which leads to a more maximal or maximal average fitness is to descend from the peak. This involves evolving in a direction of reduced average fitness that is opposed by stabilizing selection. Hence, the population is stuck on the peak at a sub maximal average fitness. When several populations are on different sub-maximal average fitness peaks, selection between populations (interdemic selection) can act.

This population genetic description has been used extensively to explain situations which cannot be explained in terms of intrademic selection. For example, body size which may have high individual fitness, and hence is selected for within a population, can reduce the fitness of the population by causing it to achieve a sub maximal average fitness and leave it open to interdemic selection.

3.3 Formalisation in Ecology

Over the last 40 years, ecology has acquired remarkable theoretical depth. Although the extensive empirical research undertaken during this period has provided the fuel for, and substantiation of, the theoretical advances, it is the vigorous employment of mathematical modelling that is largely responsible for its theoretical richness and the uncovering of complex mechanisms underlying an initially bewildering complexity. Ecology is a large domain of empirical investigation, which makes a comprehensive formalisation exceedingly complex. A comprehensive formalisation is in principle possible but is seldom employed in practice. Different aspects of ecological dynamics are usually formalised, and employed, separately since this provides the most useful framework for exploring the dynamics and directing empirical research. I provide, as an illustration, a sketch of a couple of the foundational features of ecological modelling.

Population Growth

It is an elementary task to describe mathematically *unconstrained* population growth. This, of course, was the starting point for the Malthusian claim that, unconstrained, a population will grow exponentially and for Darwin's application of it to evolutionary dynamics.

Let r be the net population growth rate (i.e., Births minus deaths in a time step) and let n_t be the tth time step (generation, although that term lacks the required precision), then:

$$n_{t+1} = rn_t$$

If one assumes, simplistically, that the growth rate r is constant, then:

$$dn/dt = rn$$

The solution to this differential equation is:

$$n_{t+1} = n_0 e^{rt}$$

If r is greater than 0, the population grows exponentially. If one assumes r is less than 0, the population decreases exponentially to 0. If one assumes $r = 0$, the population is static at its n_0 level.

This, of course is too unconstrained. A key element of Malthus' and Darwin's models is population density. For example, at some point the population outstrips its resource supply and physical space to inhabit. To take this into account r cannot be constant. A simple, but as it turns out empirically rich way to take density effects into account is to make r a decreasing function of population density. Hence:

$$r_{t+1} = r_t(1 - n/S)$$

S is the optimal, or stable, population size. Some refer to this as the carry capacity of the environment for the population. When the population density exceeds S, the growth rate, r_{t+1}, will be negative. When the population density is below S, the growth rate will be positive. When the population size is S, the population will be in a stable equilibrium and r_{t+1} will equal 0.

The differential equation for growth then becomes:

$$dn/dt = r(1 - n/S)n$$

This is the powerful *Pearl-Verhulst* or logistic equation whose solution is:

$$n_{t+1} = S/(1 + (S/n_0 - 1)e^{-rt})$$

A population obeying the logistic equation will oscillate from n_0 around S, eventually settling at the stable equilibrium S.

To this basic framework of growth one can add other features that increase the complexity of the model and its faithfulness to empirical phenomena. In

the next subsection, I show how the age structure of a population can be incorporated. Ecologists have added numerous other parameters producing a robust mathematical description of the ecological dynamics of population: competition, predator-prey interactions, foraging, resource management, frequency dependent selection, inclusive fitness, sex ratio impacts, sexual selection, etc.[46] In addition, certain assumptions underlying the description of growth in terms of the logistic equation above will not be true of some populations. In particular, the assumption that growth rate r changes linearly and the assumption that population growth (or decline) is continuous will not be true. Assuming that r changes non-linearly increases the mathematical complexity (and requires special techniques to solve the equations) but will more accurately describe many populations. Also, discrete-time models, rather than continuous-time models, will more accurately reflect many populations.

Including Age Stratification

Birth and death rates, the basis for determining r, may depend on age. For these populations, age-specific fecundities and probabilities of survival are important. To sketch how this parameter (age) can be incorporated, I assume, for simplicity, that for each age category fecundity and probability of survival are constant (discarding this assumption simply increases the mathematical complexity)[47].

Consider a population which breeds once a year and in which the birth-rate is heavily dependent on the number of breeding females. A census is taken each year prior to breeding.

Let:

$n_x(t)$	=	the number of females of age x in year t (x ranging from 0 to ω)
P_x	=	probability that a female aged x in year t survives until year $t+1$
m_x	=	average number of female offspring produced by a female aged x
f_x	=	number of these offspring surviving to age 1

Then:

$f_x = P_0 m_x$, where P_0 is the probability of a newborn (aged 0) will survive to age 1

If $x > 1$, then:

$$n_x(t+1) = P_{x-1} n_{x-1}(t)$$

If $x = 1$, then:

$$n_1(t+1) = f_1 n_1(t) + f_2 n_2(t) + \ldots + f_\omega n_\omega(t)$$

Write $\mathbf{n}(t)$ for the vector $\{n_1(t), n_2(t), \ldots, n_\omega(t)\}$, then the two equations can be written in matrix form as: $\mathbf{n}(t+1) = \mathbf{L}\mathbf{n}(t)$

[46] For a superb development of these aspects of ecological dynamics, see [Bulmer, 1994].

[47] This sketch follows Bulmer [1994]. Those wishing further details should consult his work. See also, [Vandermeer and Goldberg, 2003].

where:

$$\mathbf{L} = \begin{array}{cccccc} f_1 & f_2 & f_3 & \dots & f_{\omega-1} & f_\omega \\ P_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & P_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & P_2 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & P_{\omega-1} & 0 \end{array}$$

The first row contains net fecundities, the diagonal of P_xs contains survival probabilities. This is known as the *Leslie matrix* recognizing its promoter in ecology, P.H. Leslie, and is an age-classified model growth rate. If some other factor is more important than age (e.g., size) a more complex matrix can be constructed.[48]

The above provides an elementary glimpse at the rich mathematical apparatus employed in formalising the dynamics of population ecology systems. There is a wealth of literature on mathematical models in ecology. An excellent starting place is John Maynard Smith's little book *Models in Ecology* [1974].

4 REVISITING MODELS AND THEORIES

As the title of Section I indicates ("The Path to a Galilean Conception of Scientific Theories"), I consider scientific theories to be mathematical models of empirical phenomena and I have touched on this issue in Section I.4 above. Not all philosophers of science agree with equating theories with models *simpliciter*. For many of these philosophers, mathematical models are seen as iconic representations of the dynamics of phenomena. Theories, on the other hand, describe those representations. That is, theories provide an account of the dynamics of phenomena whereas models *merely* represent them. Hence, the syntactic account of theories, by identifying nomically necessary regularities and the logical interrelationships among them, purports to be providing an *account* of the dynamics; mathematical models, on the other hand, merely *represent* them. There are two reasons to reject this position and accept that models are theories.

First, representation in the case of mathematical models is an account of the dynamics. The language of the account is mathematics; what is said in that language (the content) embodies all of the concepts, interconnections, possible causal relations, etc. that are required to *account* for the phenomena. To hold that a syntactic account of theories is a view that specifies what it is to actually account for the dynamics, not merely represent them, is to privilege first-order predicate logic among all the available mathematical frameworks. In effect, to declare that that mathematical framework provides *an account* and all others provide *mere representations*. This seems philosophically parochial and untenable. Sometimes

[48] See, for example, Caswell [1989].

first-order predicate logic may turn out to be the most appropriate mathematical framework but in the vast majority of cases it will not.

Some philosophers have taken the position that a first-order predicate logic account of the dynamics of a system can always be provided for an account in another mathematical framework. This is an abstract possibility. However, that one *can* translate from one framework to another, and hence into the framework of mathematical logic, does not entail that one *must* do so; nor does it entail that there is an *added value* to doing so — indeed, experience suggests most often something is lost.

A second, and more decisive, reason for accepting the theory-model equivalence centres on the purposes of theorising. Among the most important purposes of developing theories are: integration of knowledge, determination/specification of an appropriate ontology, make possible explanation and prediction of events, make possible the generation of new knowledge about causal relations, entities, etc. that could not, or has not been, obtained through empirical investigation. Mathematical models satisfy all of these purposes.

However, if one still insists that theories are linguistic structures and first-order predicate logic is the natural formal language of linguistic structures, at least the forgoing provides a clear sense in which the wider context of formalisation in science — the context employing a wide and diverse array of mathematical domains — is rich, varied and powerful.

BIBLIOGRAPHY

[Amit, 1989] D. J. Amit. *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge: Cambridge University Press, 1989.

[Auyang, 1998] S. Y. Auyang, *Foundations of Complex-system Theories in Economics, Evolutionary Biology, and Statistical Physics* Cambridge: Cambridge University Press, 1998.

[Beatty, 1980a] J. Beatty. Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology. *Philosophy of Science* 47:532-561, 1980.

[Beatty, 1980b] J. Beatty. What's Wrong with the Received View of Evolutionary Theory? In P.D. Asquith and R.N. Giere (eds.) *PSA 1980*, vol. 2, East Lansing: Philosophy of Science Association, 1980.

[Beatty, 1987] J. Beatty. On Behalf of the Semantic View, *Biology and Philosophy* 2: 17-23, 1987.

[Beckner, 1959] M. Beckner. *The Biological Way of Thought* New York: Columbia University Press, 1959.

[Beth, 1948] E. Beth. *Natuurphilosophie*, Gorinchem: Noorduyn, 1948.

[Beth, 1949] E. Beth. Towards an Up-to-Date Philosophy of the Natural Sciences, *Methodos* 1: 178-185, 1949.

[Beth, 1961] E. Beth. Semantics of Physical Theories, in H. Freudenthal, pp. 48-51, 1961.

[Beth, 1965] E. Beth. *The Foundations of Mathematics: A Study in the Philosophy of Science*, Amsterdam: North-Holland Publishing Company, 1965.

[Boyer, 1963] S. H. Boyer. *Papers on Human Genetics* Englewood Cliff: Prentice Hall, 1963.

[Bulmer, 1994] M. Bulmer. *Theoretical Evolutionary Ecology* Sunderland, MA: Sinauer Associates Publishers, 1994.

[Castle, 1903] W. E. Castle. The Laws of Heredity of Galton and Mendel, and Some Laws Governing Race Improvement by Selection, *Proceedings of the American Academy of Arts and Sciences*, 39: 223-242, 1903.

[Caswel, 1989] H. Caswell. *Matrix Population Models* Sunderland, MA: Sinauer, 1989.

[Drazin, 1992] P. G. Drazin. *Nonlinear Systems* Cambridge: Cambridge University Press, 1992.
[Fisher, 1930] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press, 1930.
[Freudenthal, 1961] H. Freudenthal, ed. *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, Dordrecht: Reidel, 1961.
[Galileo, 1623] G. Galileo. *Il saggiatori*, 1623. Reprinted in English translation in C.D. O'Malley and S. Drake (eds.) *Controversy on the Comets of 1618*, Philadelphia:1960.
[Gaylord and Nishidate, 1996] R. J. Gaylord and K. Nishidate. *Modeling Nature: Cellular Automata Simulations with Mathematica*, New York: Springer-Verlag, 1996.
[Goudge, 1961] T. A. Goudge. *The Ascent of Life* Toronto: University of Toronto Press, 1961.
[Hardy, 1908] G. H. Hardy. Mendelian Proportions in a Mixed Population, *Science*, 28: 49-50, 1908.
[Hartl and Clark, 1997] D. L. Hartl and A. G. Clark. *Principles of Population Genetics* (3^{rd} ed.) Sunderland, MA: Sinauer Associates, 1997.
[Hartl, 2000] D. L. Hartl. *A Primer of Population Genetics* (3^{rd} ed.) Sunderland, MA: Sinauer Associates, 2000.
[Hull, 1974] D. L. Hull. *Philosophy of Biological Science* Englewood Cliff, N.J.: Prentice-Hall, 1974.
[Kellert, 1993] S. H. Kellert. *In the Wake of Chaos* Chicago: The University of Chicago Press, 1993.
[Landau, 1944] L. Landau. On the Problem of Turbulence, *C.R. Academy of Science*, URSS. 44:311-316, 1944.
[Langton, 1989] C. G. Langton, ed. *Artificial Life* New York:Addison-Wesley, 1989.
[Lloyd, 1984] E. Lloyd. A Semantic Approach to the Structure of Population Genetics, *Philosophy of Science* 51:242-264, 1984.
[Lloyd, 1986] E. Lloyd. Thinking about Models in Evolutionary Theory, *Philosophica* 37:87-100, 1986.
[Lloyd, 1987] E. Lloyd. Confirmation of Ecological and Evolutionary Models, *Biology and Philosophy* 2: 277-293, 1987.
[Lorenz, 1963] E. Lorenz. Determinisitic Nonperiodic Flow, *Journal of the Atmospheric Sciences* 20: 130-141, 1963.
[Luctoote and Mercier, 1998] G. Lucotte and G. Mercier. Distribution of CCR5. Gene 32-bp deletion in Europe, *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology* 19:174-177, 1998.
[Maurer, 1999] Maurer, B.A. *Untangling Ecological Complexity: The Macroscopic Perspective*. Chicago: University of Chicago Press, 1999.
[Maynard Smith, 1974] J. Maynard Smith. *Models in Ecology* Cambridge: Cambridge University Press, 1974.
[Maynard Smith, 1982] J. Maynard Smith. *Evolution and the Theory of Games* Cambridge: Cambridge University Press, 1982.
[Maynard Smith, 1983] J. Maynard Smith. Game Theory and the Evolution of Cooperation, in D.S. Bendall (ed.) *Evolution from Molecules to Man* Cambridge: Cambridge University Press, pp.445-456, 1983.
[Murray, 1993] J. D. Murray. *Mathematical Biology* (2^{nd} ed.) New York: Springer-Verlag, 1993.
[Nagel and Newman, 1958] E. Nagel and J. R. Newman. *Godel's Proof*. Routledge and Kegan Paul, 1958.
[Pearson, 1892] K. Pearson. *The Grammar of Science* London: Walter Scott, 1892. Two further editions in 1900 and 1911 published by A. & C. Black. An Everyman edition was published in 1937.
[Pearson, 1900] K. Pearson. On the Criterion that a Given System of Deviations from the Probable in the Case of Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling, *Philosophical Magazine*, **50**, 157-175, 1900.
[Provind, 1986a] W. B. Provine. *Sewall Wright and Evolutionary Biology*. Chicago: University of Chicago Press, 1986.
[Provine, 1986] W. B. Provine, ed. *Evolution: Selected Papers of Sewell Wright*. Chicago: University of Chicago Press, 1986.
[Renshaw, 1991] E. Renshaw. *Modelling Biological Populations in Space and Time*. Cambridge: Cambridge University Press, 1991.

[Rosenberg, 1985] A. Rosenberg. *The Structure of Biological Science* Cambridge: Cambridge University Press, 1985.
[Rosenberg and Williams, 1985] A. Rosenberg and M. Williams. Fitness in Fact and Fiction," Journal of Philosophy, 82:738-749, 1985.
[Rosenberg and Williams, 1986] A. Rosenberg and M. Williams. Discussion: Fitness and Primitive and Propensity, *Philosophy of Science* 53: 412-418, 1986.
[Ruse, 1973] M. Ruse. *The Philosophy of Biology*, London: Hutchinson & Co. Ltd, 1973.
[Ruse, 1977] M. Ruse. Is Biology Different from Physics? In R. Colodny (ed.) *Logic, Laws, and Life* Pittsburgh: Pittsburgh University Press, 1977.
[Segel, 1884] L. A. Segel. *Modeling Dynamic Phenomena in Molecular and Cellular Biology*. Cambridge: Cambridge University Press, 1884.
[Sen and Srivastava, 1900] A. Sen and M. Srivastava. *Regression Analysis: Theory, Methods, and Applications* New York: Springer-Verlag, 1900.
[Schaffner, 1969] K. F. Schaffner. Correspondence Rules. *Philosophy of Science* 36: 280-290, 1969.
[Skyrms, 1996] B. Skyrms. *Evolution of the Social Contract* Cambridge: Cambridge University Press, 1996.
[Skyrms, 2004] B. Skyrms. *The Stag Hunt and the Evolution of Social Structure* Cambridge: Cambridge University Press, 2004.
[Suppe, 1967] F. Suppe. *On the Meaning and Use of Models in Mathematics and the Exact Sciences*. Ann Arbor: University Microfilms International (Ph.D. Dissertation), 1967.
[Suppe, 1972a] F. Suppe. Theories, their Formulations, and the Operational Imperative, *Synthese* 25: 129-164, 1972.
[Suppe, 1972b] F. Suppe. What's Wrong with the Received View on the Structure of Scientific Theories? *Philosophy of Science* 39:1-19, 1972.
[Suppe, 1974] F. Suppe. Theories and Phenomena, In W. Leinfellner, and E. Kohler, (eds.), *Development in the Methodology of Social Science*. Dordrecht: Reidel, pp. 45-91, 1974.
[Suppe, 1976] F. Suppe. Theoretical Laws. In M. Prezlecki, K. Szaniawski and R. Wojcicki, *Formal Method in the Methodology of Empirical Science*, Wroclaw: Ossolineum, 1976.
[Suppe, 1979a] F. Suppe. *The Structure of Scientific Theories* 2nd ed. Urbana: The University of Illinois Press, 1979.
[Suppe, 1979b] F. Suppe. Theory Structure. In P.D. Asquith and H.E. Kyburg, Jr. (eds.) *Current Research in the Philosophy of Science* East Lansing: Philosophy of Science Association, 1979.
[Suppe, 1989] F. Suppe. *The Semantic Conception of Theories and Scientific Realism*. Urbana: University of Illinois Press, 1989.
[Suppes, 1957] P. Suppes. *Introduction to Logic*. Princeton: Van Nostrand, 1957.
[Suppes, 1960] P. Suppes. *Axiomatic Set Theory*. Princeton: Van Nostrand, 1960.
[Suppes, 1961] P. Suppes. A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences, in Hans Freudenthal (ed.) (1961) pp. 163-177, 1961.
[Suppes, 1962] P. Suppes, Models of Data. In E. Nagel, P. Suppes and A. Tarski (eds.) *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress* Stanford: Stanford University Press, pp. 252-261, 1962.
[Suppes, 1968] P. Suppes. The Desirability of Formalization in Science, *Journal of Philosophy* 65: 651-664, 1968.
[Tarski, 1944] A. Tarski. The Semantic Conception of Truth and the Foundations of Semantics, *Philosophy and Phenomenological Research* 4: 341-376, 1944.
[Tarski, 1946] A. Tarski. *Introduction to Logic*, New York: Oxford University Press, 1946.
[Tarski, 1953] A. Tarski. *Undecidable Theories*, Amsterdam: North-Holland Publishing Company, 1953.
[Tarski, 1956] A. Tarski. *Logic, Semantics and Metamathematics*, Oxford: Oxford University Press, 1956.
[Thompson, 1983] P. Thompson, The Structure of Evolutionary Theory: A Semantic Approach, *Studies in History and Philosophy of Science* 14: 215-229, 1983.
[Thompson, 1985] P. Thompson. Sociobiological Explanation and the Testability of Sociobiological Theory, in J.H. Fetzer (ed.) *Sociobiology and Epistemology* Dordrecht: D. Reidel, pp. 201-215, 1985.
[Thompson, 1986] P. Thompson. The Interaction of Theories and the Semantic Conception of Evolutionary Theory, *Philosophica* 37: 73-86, 1986.

[Thompson, 1987] P. Thompson. A Defence of the Semantic Conception of Evolutionary Theory, *Biology and Philosophy* 2: 26-32, 1987.
[Thompson, 1988a] P. Thompson. The Conceptual Role of Intelligence in Human Sociobiology. In H.J. Jerison and I.L. Jerison (eds.) *Intelligence and Evolutionary Biology* New York: Springer-Verlag, 1988.
[Thompson, 1988b] P. Thompson. Logical and Epistemological Aspects of the 'New' Evolutionary Epistemology, *Canadian Journal of Philosophy* 14 (supplementary):235-253, 1988.
[Thompson, 1989] P. Thompson. *The Structure of Biological Theories*. Albany: State University of New York Press, 1989.
[Thompson, 1992] P. Thompson. Mathematics in the Biological Sciences, *International Studies in the Philosophy of Science* 6: 241-248, 1992.
[Thompson, 1995] P. Thompson. Managing Complexity and Dynamics: Is There a Difference Between Physics and Biology, *Canadian Journal of Philosophy* 20 (supplementary volume) 275-302, 1995.
[Vandermeer and Goldberg, 2003] J. H. Vandermeer and D. E. Goldberg. *Population Ecology: First Principles* Princeton: Princeton University Press, 2003.
[van Fraassen, 1970] B. C. van Fraassen. On the Extension of Beth's Semantics of Physical Theories, *Philosophy of Science*, 37: 325, 1970.
[van Fraassen, 1972] B. C. van Fraassen. A Formal Approach to Philosophy of Science. In R.E. Colodny (ed.) *Paradigms and Paradoxes*, Pittsburgh: The University of Pittsburgh Press, 1972.
[van Fraassen, 1980] B. C. van Fraassen. *The Scientific Image*. New York: Oxford University Press, 1980.
[van Fraassen, 1981] B. C. van Fraassen. Theory Construction and Experiment: An Empiricist View. In P.D. Asquith and R.N. Giere (eds.) *PSA 1980* vol. 2, East Lansing: Philosophy of Science Association, pp. 663-677, 1981.
[Waddington, 1968–72] C. H. Waddington, ed. *Towards a Theoretical Biology* (4 volumes) Edinburgh: Edinburgh University Press, 1968–72.
[Weinberg, 1908] W. Weinberg. Uber den Nachweis der Vererbung Biem Menschen, *Jahreshefte des Vereins für Vaterländische Naturkunde in Württemburg*, 64: 368-382, 1908.
[Weisheipl, 1967] J. A. Weisheipl. Galileo and his precursors. In E. McMullin (ed.) *Galileo: Man of Science*. New York: Basic Books, 1967.
[Williams, 1970] M. B. Williams. Deducing the Consequences of Evolution: A Mathematical Model. *Journal of theoretical Biology* 29: 343-385, 1970.
[Williams, 1973] M. P. Williams. Falsifiable Predictions of Evolutionary Theory, *Philosophy of Science* 40: 518-537, 1973.
[Woodger, 1932] J. H. Woodger. *The Axiomatic Method in Biology* Cambridge: Cambridge University Press, 1932.
[Woodger, 1939] J. H. Woodger. *The Technique of Theory Construction* Chicago: University of Chicago press, 1939.
[Yule, 1902] G. U. Yule. Mendel's Laws and Their Probable Relation to Intra-racial Heredity, *New Phytologist*, 1: 193-207; 222-238, 1902.

This page intentionally left blank

FUNCTIONS

Tim Lewens

1 THE PROBLEM OF BIOLOGICAL FUNCTIONS

The problem of biological functions arises because biologists' use of the function concept appears distinctive in comparison with that of physicists or chemists. True enough, physicists and chemists ask questions about functions. A chemist, for example, might try to ascertain the precise function of some reagent in an inorganic process. But in articulating the function of chlorofluorocarbons (CFCs) in the breakdown of atmospheric ozone the chemist would not suggest that CFCs are for breaking down ozone, nor that she has discovered the purpose of CFCs, nor would she think that she has somehow explained why CFCs feature in this system of reactions. Compare this with a claim like 'the function of the peacock's gaudy tail is to attract mates'. Here there is, on the face of things at least, a strong connotation that peacocks' tails are for attracting mates, that the purpose of the peacock's tail is to attract mates, and that attracting mates somehow explains why peacocks have such gaudy tails. Let us call function talk that carries some or all of these connotations *heavy function talk*, and function talk of the more modest form that our chemist uses *light function talk*. In day-to-day life it is in the contexts of artefacts designed by intelligent agents, or actions performed by intelligent agents, that we tend to find this kind of heavy function talk. Hence one might harbour an initial suspicion regarding the propriety of heavy function talk in biology.

Very broadly speaking, there are four types of reaction one might have to these preliminary observations, and to the suspicions they give rise to. The first is to argue that serious biologists do not make use of heavy function talk at all; perhaps this appears in some popularisations of their work, but it is an exaggeration to think that it plays any role in the practice of the science. The second is to concede that while some biologists do use heavy function talk, these biologists are making a simple mistake, for talk of this kind has no business in a grown up science. The third is to argue that heavy function talk is justified, and even useful, but not because it is true of organic systems: rather, function talk is metaphorical, or its use relies on analogies of some kind between organic systems and other systems (most obviously artefacts or actions) where heavy function claims can be asserted truly. The fourth option is to argue that heavy function talk is justified in biology, and to argue that it can be truly asserted of biological systems, in virtue of their instantiating purely biological properties.

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Published by Elsevier B.V.

Recent philosophical work on functions has come down overwhelmingly in favour of option four ([Matthen, 1997], who favours option three, is an exception). That is surprising, because every one of these options offers some promise of achieving what is probably the major goal of modern philosophical analyses of function, namely, to come to an understanding of the appearance of function talk in biology without the need to invoke either reverse causation or divine design [Buller, 1999a, 19]. One of my main aims in this essay is to encourage further exploration of the other three options.

2 ORGANISMS, ARTEFACTS, AND AGENTS

The four options I outlined in the previous section concerned what kind of significance we should attach to function talk in biology. There is another important distinction, orthogonal to this one, that concerns how we are to analyse that talk. This distinction is between theories of biological function that take their inspiration from what I call the *artefact model* of organic function [Lewens, 2004], and theories that take their inspiration from the *agent model* of organic function. In the first case, one sees the functions of organic parts as similar in kind to the functions of the parts of artefacts. In the second case, one sees the organism as a whole as akin to an agent aiming at some goal.

Agents have goals in virtue of facts that are internal to them: what an individual is aiming at is a matter of his motivational states, his dispositions regarding what kinds of events would cause him to cease his search, or some such. (I am ignoring complications presented by externalist accounts of mental content.) Accounts of function that draw on the agent model explain the presence of heavy function talk in biology by seeking facts about the internal organisation or development of organisms, which are analogous to those facts about an agent's motivations that make it the case that the agent has a particular state of affairs as its goal. Agent-based accounts of function talk in biology consequently tend to regard *goal-directedness* as the basic notion to be analysed. *Functions* are logically secondary to goals: functions of organic parts or processes are contributions to overall organic goals.

Artefact-based accounts tend to reverse the logical relationship between the functions of the parts and processes of organisms and organisms' goals. (This says nothing one way or the other about whether they make functions more basic than goals in general.) Artefacts have their functions in virtue of facts external to them: what a tool is for depends on its history of design or its pattern of use, not on its internal organisation. Accounts of function that draw on the artefact model explain the presence of heavy function talk in biology by seeking a biological process that is external to organisms themselves, and which is analogous to the processes of design or use that ground function claims in the context of artefacts. Accounts of this sort are often illustrated using organic traits which are more naturally described as functional, rather than goal-directed. Although the eyespots on a peacock's tail may have the function of attracting mates, it seems strained to

say that attracting mates is their goal: after all, the eyespots are not responsive to the presence of mates — their structures are static. To the extent that parts of organisms, or organic systems, are regarded as having goals this also rests ultimately, for artefact-based theories, on biological processes external to organisms. According to these views, just as a thermostat has the goal of regulating heat in virtue of its design history, so homeostatic systems aim at maintaining organic equilibrium in virtue of their histories.

The artefact/agent distinction is not intended to be exclusive. One could, for example, fashion a theory of functions around the conception of organisms as agents whose traits are their tools. Another conception, encouraged by the widely-used distinction between replicators and vehicles, is of genes as agents, and of organisms as the tools used by genes to enable their replication. One might wonder whether all theorists of biological function who adopt some variant of the 'artefact model' for the functions of organisms and their parts must eventually find some biological analogue to an agent. After all it is plausible to think that in the domain of artefacts, the external facts which fix the functions of tools ultimately depend on intentions, which in turn depend on the goals of agents. This may explain, in part, why Daniel Dennett [1995] sometimes casts the role of the evolutionary biologist as that of 'reading Mother Nature's mind': he conceives of natural selection in the abstract as an agent, and of organic traits as tools designed by that agent. The job of the biologist, says Dennett, is to discover (metaphorically speaking) what Mother Nature had in mind in building organisms one way rather than another. I do not propose any solution to the question of whether goals are always conceptually prior to functions here. The role for the artefact/agent distinction is merely to explain some of the initial directions in which philosophical analyses of the function concept proceed. To recap, some theorists look to organisms' developmental and regulatory processes to justify a view of the organism as akin to a goal-directed agent, while others look to evolutionary processes to justify a view of the organism as akin to a collection of designed tools.

The artefact/agent distinction is orthogonal to the fourfold distinction I introduced near the beginning of this essay because, to take just one example, regardless of whether we think function talk in biology is literally true of biological systems, or merely expresses a series of fruitful metaphors, we would still have to decide whether that talk should be analysed and explained in ways that relate organisms to goal-directed systems, or to designed systems, or to both. Almost all contributors to the debate over the past twenty years or so have adopted an artefact-based model of function. Once again, a central aim in this essay is to explore the prospects for agent-based models, and to encourage others to give these accounts a second chance.

There is a long tradition in British biology in particular that focuses on the similarities between organisms and artefacts, and it is surely this tradition that explains why artefact-based accounts of function are so prevalent in philosophy of biology today. This tradition can be observed most clearly in the work of Natural Theologians like William Paley [Paley, 1802]: for these writers organisms were

artefacts, and God was their designer. Richard Dawkins calls one of his books *The Blind Watchmaker* [Dawkins, 1986] and in giving the book this title he makes it clear that although he is opposed to Paley's view that only intentional design can explain the makeup of the natural world, he is not opposed to characterising natural objects as akin to watches. Natural selection is blind, but the end result is an organism whose parts are all well-designed for their roles. Similar appeal to the artefact model, particularly in the conception of natural selection as an ersatz designer, can be seen in Krebs and Davies' rationale for attending to adaptive 'problems' in behavioural ecology:

> Visitors from another planet would find it easier to discover how an artificial object, such as a car, works if they first knew what it was for. In the same way, physiologists are better able to analyse the mechanisms underlying behaviour once they appreciate the selective pressures which have influenced its function. [Krebs and Davies, 1997, 15]

The tradition that flows from natural theology is what establishes the features of biological discourse that a philosophical account of functions then seeks to preserve; in the context of a different tradition, the desiderata of a theory of functions can be different too.

If we begin by noting the differences between organisms and artefacts, rather than their similarities, then we are likely to be tempted to construct a theory of functions of the form suggested by the agent model. Kant and Aristotle, for example, take pains to point out how unlike artefacts organisms are [Ginsborg, 2004]. Kant points out that an artefact is something that suggests to us a rationally determined organisation, but when we see an organism, we are not merely put in mind of a rationally ordered object. In Kant's terms, an organism is also, in three senses, 'both cause and effect of itself' [Kant, 1952, 18], and in these respects organisms are unlike artefacts. Kant sees organisms as self-reproducing (trees come from trees), self-realising (a tree grows by drawing in and synthesising its own supply of nutrients), and self-dependent (each part of the tree depends for its continued existence on the contributions of the other parts): in short, Kant says, an organism is a 'self-organised being' [*ibid.*, 22].

Whatever one thinks of the precise way in which Kant spells out these disanalogies, it is clear that we observe very different modes of generation, maintenance and repair in the natural and artificial worlds. What makes a whale is another pair of whales; what makes a watch is a watchmaker. Watchmaking is a very delicate business; the whole project may stall if a part is missing, or if the environment of manufacture is disturbed. Whalemaking, on the other hand, is fairly resilient to changes in developmental conditions. A whale is a self-maintaining entity; if its environmental conditions change, then internal adjustments (of heart rate, for example), can compensate for those changes and ensure an ongoing organic integrity. Watches, on the other hand, are not so resilient, nor so resourceful. Divers' watches might sometimes remain in working order across changes in temperature and pres-

sure, but not because they make compensating internal adjustments. Whales are also self-repairing; if it is injured, a whale can grow new tissue, for example. But watches cannot repair themselves in this way. This way of thinking of organisms views them not so much as bags of good tools, but instead as analogous to people who are striving towards some goal. The organism is an object whose goal is to attain and then to maintain its integrity (or perhaps its capacity to reproduce) in the face of a changing environment. A person, too, will change their behaviour to get what they aim at if the environment is changing. If you are trying to lose weight, you will find a new gym when the one you have been going to closes down. And if a person's plans are blown off course (their crash diet repays them with a swift weight gain) then, so long as they still have their goal, they will pick themselves up and set off after their goal again. If we stress the differences between whales and watches, and instead focus on the similarities between whales and weight watchers, we are likely to arrive at an *agent-based model* of heavy function talk.

It is no coincidence, then, that if one is fed on a diet of examples from modern Anglo-American evolutionary biology one is likely to construct an artefact-based model of function, while if one is instead fed on a diet of developmental biology one is more likely to construct an agent-based model of function. The study of development makes goal-directedness the most tempting focus for an account of function. The study of evolution instead encourages us to conceptualise organisms as sets of tools for solving environmental problems. Thus George Williams, an influential figure among contemporary adaptationists, makes clear allusion to the artefact model when he begins his celebrated work *Adaptation and Natural Selection* with the words, 'Evolutionary adaptation is a special and onerous concept that should not be used unnecessarily, and an effect should not be called a function unless it is clearly produced by design and not by chance' [Williams, 1966, vii]. It is interesting to contrast this with some introductory remarks in C. H. Waddington's *The Strategy of the Genes*, a work that focuses strongly on development. Here Waddington does not focus on adaptation to environment as characterising the organic realm, but rather he mentions both the dynamic manner in which the form of an individual organism is maintained, and:

> ... the second major peculiarity of living things, their 'directiveness' ... This refers to the fact that most of the activities of a living organism are of such a kind that they tend to produce a certain characteristic end result. The most inclusive type of end-result, it is often stated, is the continued life of the organism; actually its reproduction, and the passing of its hereditary qualities to the next generation, should be regarded as a still more general goal. [Waddington 1957, 2]

More recently, Stuart Kauffman, known for work on the 'self-organisation' of complex organic systems, makes direct reference to a Kantian view of goal-directedness when describing 'autocatalytic sets'. These are sets of molecules that

act as mutual catalysts and reagents, which Kauffman views as primitive living systems:

> In a collective autocatalytic set, there is no central directing agency. There is no separate genome, no DNA. There is a collective molecular autopoetic system that Kant might have been heartened to behold. The parts exist for and by means of the whole; the whole exists for and by means of the parts. [Kauffman, 1995, 275]

3 THE SELECTED EFFECTS ACCOUNT

In this section I want to evaluate the most influential modern theory of functions. This is the Selected Effects, or SE account. The SE account has its roots in Wright's [1973; 1976] account of functions, and has been developed and defended in most detail by Ruth Millikan (e.g. [Millikan, 1984; 1989]) and Karen Neander (e.g. [Neander, 1991a, b]). SE analyses of one sort or another dominate two recent collections on functions ([Buller, 1999b] and [Amundson *et al.*, 1998]). The SE account gets much of its appeal from the way it seeks to supply a biological surrogate for the natural theologians' intelligent design. As the quotations from Williams, and from Krebs and Davies indicate, many modern evolutionary biologists continue to see the traits of organisms as useful tools produced by design in response to problems laid down by the environment. Elsewhere, Williams also licenses an updated version of Paley's argument from design: he explicitly recommends that biologists take note of Paley's criteria for determining when a trait should be attributed to design rather than chance — when, that is, it is an adaptation [Williams, 1992, 190]. Of course such biologists do not think that an intelligent designer is responsible for these functional traits; however, if we begin by supposing that in the realm of artefacts, functions are established through the intentions of users or designers, then one who is liable to describe organisms in terms of function and design needs to find a non-intelligent natural process that can play the role of intelligent design in the organic realm. Because the whole range of artefact talk is found in biology, and because it tends to be used in just the same way that it is used on its home turf, we want a way of making sense of biological functions that respects the connotations of heavy function statements when they are applied to artefacts.

Philosophers typically find three such connotations when heavy function talk is used, regardless of whether we are dealing with organisms or artefacts:

> *Function statements are explanatory:* Suppose we are looking at the radiator at the front end of a car. The questions 'What is that for?', 'What is that thing's function?' and 'Why is that thing there?' all seem equivalent in this context. When we give an answer that states a function — 'It's for keeping the engine cool' — we therefore answer the question 'Why is it there?', and we thereby explain the presence

of the radiator. In short, function statements explain the presence of functionally characterised artefacts. Function statements also explain the presence of functionally characterised traits. When we say that the function of the sail on the back of Dimetrodon is to regulate heat, we also give an answer to the question 'Why is the sail there?', and we thereby explain the presence of the sail.

Functions can be distinguished from accidents: Artefacts have all kinds of useful effects that are not their functions. These have come to be known as 'accidents', or accidental benefits. A bible kept in a breast pocket might have the beneficial effect of deflecting an otherwise deadly bullet, but that is not what the bible is for. The bible is for spreading the word — that is its function, and the preservation of life is only an accident. Once again, what goes for artefacts goes for organisms. Organic traits may have plenty of beneficial effects that are not their functions. The standard example is that of heart sounds: human hearts make noises, which can often benefit their bearers by assisting doctors in making diagnoses that lead to subsequent medical treatment. But making diagnostic noises, although an effect of hearts, is not what they are for. Hearts are for pumping blood.

Functions statements are normative: Function statements do not describe how an artefact is behaving; rather, they describe how it should behave. A CD player that repeatedly skips has the function of playing music, even though it cannot play music. It makes sense to say that the CD player has gone wrong, or that it is malfunctioning, because the CD player continues to have a function even though it cannot perform that function. Similarly, to say that the function of the eye is to assist in vision is to say what eyes ought to do, even though some eyes may have this function and be unable to perform it on account of being clouded by cataracts.

It is largely because of a conviction that any account of biological functions must justify these connotations of function statements by reference to biological facts alone, that many philosophers have rejected what has become known as the Causal Role (CR) account of functions. The Causal Role account, which we owe to Robert Cummins [1975], vies with the SE account for attention in most current work on functions. On Cummins' view, function attributions feature in *functional analyses* of complex systems. Functional analyses explain the capacities of a system (whether that be a washing machine, a car, the mammalian reproductive system or the Belgian economy) by reference to the contributions of the system's heterogeneous parts. The CR account says that an item's functions are the contributions that it makes towards a capacity of a containing system. The questions of which containing system, and which capacity, are the relevant ones by which to attribute functions are decided by the interests of the investigator. So on this view, the function of the heart is to pump blood because, in the context of the capacity

of the circulatory system to bring nutrients to cells, the heart's contribution is pumping.

The CR account, at least the entry-level version, has trouble making sense of all three of the connotations of function ascription listed above (although see [Davies, 2001], for an attempt to defend an augmented CR account). First, CR functions are explanatory, but they give the wrong kind of explanation. If the function of a token heart is one of its effects, then while that explains the capacities of the systems that contain the heart, it does not explain the presence of the token heart itself.[1] Second, the CR account cannot make sense of the function/accident distinction. In the context of the containing system of hospital treatment practices, the heart does have the function of making diagnostic sounds. Third, the CR account cannot make sense of malfunction; if having a function F demands having effect F, then malfunctions are impossible, for objects malfunction when they have function F yet are unable to yield effect F.

We can add a fourth problem to this list, the problem of promiscuity. Remember that biology seems to be quite distinctive in its use of function talk, so whatever analysis we give of biological functions should explain why this vocabulary is used in biology in a way that does not feature in the physical science. But it seems hard to restrict CR functions to biology: any item — whether organic or inorganic — that makes a contribution to some capacity of a containing system will have a CR function. To return to our opening example, in the context of an explanation of the breakdown of atmospheric ozone, CFCs will be credited with a function. But chemists do not use heavy function talk. So if heavy function talk is what we are trying to understand, the CR account does not explain its unique role in biology. In fact, most current writers on functions are pluralists [Godfrey-Smith, 1993]: they think that some, but not all, biological function talk is heavy function talk. Sometimes when biologists speak of functions they are doing nothing more than talking of the role of some item in a containing system. Functional analyses of these kinds seem to occur in physiology and biomechanics. This is light function talk. The CR account seems well placed to make sense of light function talk, but when biological function talk is heavy, the CR account is inadequate.

When we look to artefacts, it seems that we can explain our three connotations by appeal to the simple view that the function of an object is just whatever its maker (or perhaps its user) intended it to do. If function claims for artefacts are claims about their intended effects, then we can see why function claims tell us why an object is there (they tell us the intention that led to its appearance), we can see why functions need to be distinguished from accidents (not every beneficial effect will have been intended), and we can see why functions express norms (an object that cannot do what it is intended to do is malfunctioning).

The selected effects account takes the bare bones of this approach — functions

[1]Note that Cummins is well aware of this feature of his view. It does not bother him, because he denies that the presence of any token heart is genuinely explained by the pumping of blood. In other words, he doubts that the first connotation of heavy function talk is one a good theory should seek to justify.

are effects of objects that do not owe themselves to chance, and which are bestowed by some external process — and adapts them to the biological realm. Natural selection takes the role of a designer. If the function of an artefact is what a designer or user intended it to do, then it is only a short step to say that the function of an organic trait is whatever it was *selected for*. To say that some trait T was selected for F is to say that T increased its frequency in a population over some alternative because it had effect F [Sober, 1984, 100]. To say that the dark pigment of the wings of the peppered moth was selected for camouflage is to say that moths with dark wings increased their frequency over moths with lighter wings because the dark wings provided camouflage. So now we can see how the SE account makes sense of our three connotations of function ascription, and how it does it in a way that closely follows how we grounded artefact functions. According to the SE account, function claims explain the presence of the functionally characterised item because they are, in spite of their present-oriented appearance, historical claims about the selective history of the trait. Functions are distinguished from accidents because some of the fitness-enhancing effects of a trait type will be too new to explain the current frequency of that trait type in the population. And the reference to natural selection enables us to ground the normativity of function statements, because a trait can be a token of type whose presence is explained by selection for some effect, even though the trait in question does not have that effect. That is how a token heart that is unable to pump blood can still have pumping blood as its function. Finally, the SE account does not face the promiscuity objection levelled at CR functions. Since the SE account makes functions dependent on selection, the SE account restricts functions to systems that undergo selection, and hence explains why biologists, but not physicists or chemists, make use of this kind of talk.

One further point of clarification is worth making here. In order to cut off worries that by looking to a trait's history of selection, the SE account makes the loss of function impossible, most theorists these days also stipulate that the selective regime that determines a trait's function is that which occurs in the recent history of a trait. The most widely adopted theory of functions these days is some variant of the 'modern history' account [Griffiths, 1993; Godfrey-Smith, 1994].

To illustrate a last strength of the SE account I want to turn back to Kant. Kant worries that a purely mechanistic explanation of organic form would always leave too much of an air of 'contingency' to that form:

> So where the structure of a bird, for instance, the hollow formation of its bones, the position of its wings for producing motion and of its tail for steering, are cited, we are told that all this is in the highest degree contingent if we simply look to the *nexus effectivus* in nature, and do not call in aid a special kind of causality, namely, that of ends (*nexus finalis*). [Kant, 1952, 4]

Consider a house. One could give a mechanical explanation for how the house comes to be built by citing the positions of all the bricks and window frames

around the building site, and then cataloguing how various Newtonian forces act on them so that they fall into their final resting places. However, this does not explain in a satisfactory way why, when the bricks and window frames could have come together into a useless pile, they did in fact form such an elegant dwelling. Here we point to a rational plan that explains this ordering, and Kant is unable to see how a mechanistic explanation of development could escape a similar charge of contingency unless supplemented with appeal to some type of rational plan, too. Darwin's great breakthrough is to show, through natural selection, why organisms come to be well ordered. Selection explanations show us how the fact that a trait of a particular form *would* augment an organism's reproductive fitness, can bring about the existence of a trait with just such a form. The creativity of natural selection is an important factor in undermining Kant's worry that mechanical explanations will always leave contingent the adaptedness of organic form.

4 PROBLEMS WITH THE SELECTED EFFECTS ACCOUNT

The SE account has generated a very large literature in recent years, and there is no sense in trying to survey every objection that has been raised here ([Buller, 1999b] and [Allen *et al.*, 1998] are comprehensive collections of important recent articles). Most objections turn on two features of the theory, both of which yield somewhat quirky results. First, the SE account is backward-looking. Traits have functions in virtue of their selection history, and for this reason the SE account is sometimes called an *etiological* account. Second, the account is highly theory-laden. Biological functions are analysed in the technical terms of modern evolutionary theory. The first feature commits the SE account to the following two pronouncements:

> **A**: A new mutation — let's say a pair of fully functional eyes, which enhances fitness greatly — arises in a population. Organisms with eyes increase in frequency in the population. Consider the first generation after the original eyes arise. It is, strictly speaking, a mistake to say that there are lots of organisms with eyes in this generation because earlier eyes had the function of assisting in vision. The earlier eyes had no functions at all, for they had no history of selection.
>
> **B**: An atom-for-atom replica of a lion, which coalesces one day from a chance coming-together of molecules, has no organ for pumping blood even though it has an organ that does all the same things as a normal lion's heart. With no evolutionary history, no parts of this lottery lion have any functions.

The second feature commits the SE theorist instead to a third type of claim:

> **C**: It appears that Harvey correctly identified the function of the heart as that of pumping blood. But Harvey cannot, in fact, have known the correct biological function of the heart, for Harvey (pre-dating Darwin

as he does) cannot possibly have been competent in the use of the SE concept of function.

None of these problems is fatal, although they are a little awkward. As far as scenarios A and B go, the SE theorist can simply bite the bullet. Function claims are already historical explanations in disguise, which is why it is not quite right to say that eyes exist because their function is to assist in vision. And when we remember that we have to deny only that the *purpose* of the heart of an instant lion is to pump blood, not that instant lion hearts are able to pump blood, it is not so clear how worrying the SE account's pronouncements are. The instant lion is an unlikely animal indeed, and he has all the capacities and causal powers of a normal lion. In such cases the rule should be 'spoils to the victor', as David Lewis sometimes argues. The odd awkward pronouncement is a cheap price to pay for an otherwise strong theory. The cost is not so cheap if we want to use the SE theory for work in teleosemantics, for there we end up denying not merely that the mental states of instant persons have the biological purpose of representing states of affairs, we deny also that they have representational capacities at all.

Scenario C is also less damaging than many suppose. The SE theorist simply needs to make clear the scope of her analysis. She is trying to make sense of function claims in modern biology. It is perfectly legitimate to give an account that only applies to modern biology, although an additional requirement of the theory is that it should be able to explain why there is such continuity in how functions have been described in spite of a radical change in the processes thought to underpin such function talk. The suggestion that selection is at least a design-like process — a suggestion that seems to account for much of the appeal of the SE account in grounding heavy function talk — promises just such an escape route for the SE theorist who does not wish to belittle what Harvey did for us.

A more interesting objection, although again not fatal, is the claim that the SE account does not suffice for the individuation of trait types (important articles on this topic include [Matthen, 1998], [Neander, 2002] and [Griffiths, manuscript]). Sometimes it seems that SE theorists think that their account gives us the resources we need to say what makes two token traits instances of the same type. On this view, what it takes to be a heart is to have a certain function. Yet that cannot be the whole story, and the SE theorist needs only to retain her modesty here. A good way to see why there must be some way of saying when two traits are the same that does not rely on functions is to look again at the concept of malfunction. A damaged eye cannot help for seeing, and neither can an ear. Yet an ear is not malfunctioning in this respect, but a damaged eye is. The SE account must make sense of this by saying that helping to see is not the function of ears. That, in turn, is to say that helping to see does not explain why ears increased their frequency, while it does explain why eyes increased their frequency. But if two traits are pleiotropically linked, then the fitness contributions that explain the increase in frequency of one will also explain the increase in frequency of another. Suppose, then, that earlobes are present in some population of organisms because they are pleiotropically linked to an essential part of the eye — the lens, say. If assisting

in vision explains the presence of lenses, it explains the presence of earlobes, too. Why not say, then, that earlobes have helping to see as one of their functions, hence that all earlobes are malfunctioning? We need to clarify the SE account. A token trait of type T has function F just in case *past tokens of the same type T* contributed to fitness by producing effect F, and thereby explained the proliferation of tokens of type T. Earlobes don't have helping to see as a function, because the past token traits of that same type did not contribute to fitness by assisting in vision. But if we are to say this, we need some way of picking out trait types independently of how they have contributed to fitness. The obvious way to do this is by appeal to homology. So, on this view, we sort tokens into types using the concept of homology, and we then assign functions to trait tokens according to how earlier homologous tokens contributed to fitness. Any classification of traits by function is parasitic on a prior classification by homology.

5 MALFUNCTION

This discussion of homology sets up useful background for a more serious problem that faces the selected effects account. Elsewhere I have argued that the SE account is not, in fact, uniquely able to meet the desiderata usually accepted for a theory of functions [Lewens, 2004]. Rather than repeat those arguments here, I want to ask why an account of evolutionary function needs to make sense of malfunction at all. (My arguments here draw on those by [Davies, 2000].) My attack will be two-pronged. On the one hand, I want to downplay the theoretical interest in the category of malfunction from the perspective of evolutionary biology. On the other hand, I want to suggest that what the selected effects account says about malfunction is rather awkward. The result is that I am sceptical of whether the ability of an account of biological functions to ground claims about malfunction should really be viewed as a desideratum for such accounts.

Let us begin with the question of the role of malfunction in evolutionary theory. Variation is a feature of all biological populations. What is the theoretical interest in marking some of this variation as *malfunctioning*? The evolutionary biologist is primarily interested in explaining the changes of frequencies of various traits in a population. Now we can assign functions to traits of various types according to the effects that they have which account for their representation in the population as a whole. Here we are, in effect, treating the population as a system, and treating classes of resembling traits as parts of that system whose frequencies in the system stand in need of explanation. Within this type of explanatory project there is no need to posit the category of malfunction as a correlate of function: this kind of project is a Cummins-style functional analysis of a population, and we have seen that such an analysis stops short at a purely descriptive account of the systems whose capacities it explains.

Some traits that arise as single mutational oddities in a large population — whether fitter than the average or less fit — may be so rare that they can be ignored in the analysis of that population. These, however, can be viewed as

anomalies — rare instances that conform to no pattern of theoretical interest — rather than as malfunctioning traits. They are discounted for the same reason that any science might discount an extremely improbable accident — because they are too unusual to have general impact, not because they are examples of something that has isn't doing what it is supposed to. Some traits that are present at a high frequency may appear to be detrimental to the fitness of their bearers. The biologist might try to explain their presence in terms of drift, or in terms of pleiotropic linkage to some other trait that is under selection. But neither explanation describes the trait as malfunctioning. Suppose, however, that the trait is present at high frequency because it is descended from a trait that was selected for, but whose developmental pathway has been perturbed by some large-scale environmental change, with the result that what is now present is a new trait which reduces fitness. If the perturbation is short-lived — one generation in duration — we might write the whole thing off as an anomaly in the context of the overall dynamics of the population, rather than a malfunction. If, on the other hand, the perturbation is long-lived, then we will need to readjust the terms of our population analysis to see the new trait as one of lasting evolutionary interest. In neither case is it pressing to judge that the new trait is malfunctioning. All of these observations are borne out by the fact that while evolutionary biologists regularly hypothesise functions for traits, they are very rarely in the business of offering claims about trait malfunctions.

What does the SE account say about malfunction? The basic gist of the SE account is that a trait is malfunctioning just in case it fails to have the effect that explains its presence. More formally, a token trait is malfunctioning if and only if it does not have the effect (or is less proficient in its performance of the effect) that accounts for the prevalence of traits of that type. This sounds sensible. Just as an artefact malfunctions if it cannot do what it is intended to do, so a trait malfunctions if it cannot do what it is selected for. But there are cases where it is not so clear what we should say. Imagine a population of peppered moths that contains both dark and light forms in equal proportions. Suppose there has been no predation from birds. Now suppose that birds arrive in the moths' habitat, and the dark moths begin to increase their frequency because they are better camouflaged against the dark trees. It seems clear that once the dark moths are at fixation, in the event that a light moth appears in the population through mutation we can say that these light wings are malfunctioning. But what if the light moth appears in the population through migration from a population in which the trees are light in colour? Presumably the SE account will say that this moth's wings are malfunctioning too, although this is not clear, especially if there is no pedigree of providing camouflage in this moth's ancestors. What about the wings of the light coloured moths a few generations after selection for camouflage begins, and before the dark variety goes to fixation? Here is what the SE account should not say: these light coloured wings are not malfunctioning, because camouflage does not explain the presence in the population of light coloured wings, hence moths with light coloured wings do not have camouflage as a function. The SE account

cannot say this, because it equates having a function with being able to perform that function; hence, it rules out malfunction altogether. This move would entail that a mutant light coloured moth that appears in the population after the dark wings go to fixation is not malfunctioning.

So what should the SE account say about light wings while the dark ones are going to fixation? Are they malfunctioning or not? The account could say 'no' of many of these moths, perhaps on the grounds that a trait token only has a function F if token homologous traits in its ancestors contributed to fitness by F-ing. Ancestors of many of the light-winged moths did not (we can suppose) have wings that contributed to fitness in this way. Alternatively, the account could say 'yes', perhaps on the grounds that a trait token has a function F in case it is homologous with other tokens that have contributed to fitness in the past by F-ing. The light wings and the dark wings are homologous, and dark wings have contributed to fitness by providing camouflage. This second option might seem strange, but if the basic gist of the SE account is that the function of a trait is whatever effect explains the selective success of traits *of that type*, we might say that we should treat dark moth wings, light moth wings after the dark ones have gone to fixation, and light wings while the dark ones are going to fixation, all as 'traits of the same type'.

Now let me come clean: the primary point of this excursion is not to suggest that it is impossible for the SE account to manufacture a set of hard and fast conditions, which are phrased in purely biological terms, that will tell us when some trait is to count as malfunctioning. There may be all kinds of ways to achieve this job, and I have sketched some above. My point, rather, is that there is no particular way of doing this suggested by the practice of evolutionary biology itself, nor does that practice make such a project needful. The biologist who wishes to understand changes in the frequencies of traits in a population has no need to decide which should count as malfunctioning, and which should count as functioning: it is enough to construct a model that can explain trait fluctuations.

One might reply by pointing out that the concept of malfunction does play an important role in the context of assessments of health and disease: diseased traits are those that have 'gone wrong'. Suppose the environmental change in the example above, the one that results in the failure of a very high proportion of traits in some population to augment fitness in the usual way, is in fact the sudden presence in the human population of a virus that affects the development of the kidneys so that they do not filter blood efficiently. Here humans may well want to assert that most kidneys are now diseased. I tried to suggest above that the study of evolution does not, in itself, need to make assessments of health and disease in cases like this. This does not show that the health/disease distinction should not be understood in biological terms (see [Boorse, 1997], for arguments that it should be). Perhaps it is possible to equate disease with malfunction, and to define malfunction in biological terms. Whatever the result of that debate, we do not find such a malfunction concept playing a role in evolutionary biological inquiry; rather, we construct that concept from evolutionary biology's conceptual

repertoire. Hence I am sceptical of whether an account of biological function that seeks to do justice to the practice of evolutionary biology needs to respect malfunction claims, and I am equally sceptical of the credit we should give to the SE account for doing so. The idea that malfunction claims are an important part of the practice of evolutionary biology is best understood as a hangover from natural theology. Once we eliminate the need for an account of biological function to respect malfunction claims, we begin to downplay the seriousness of heavy function talk within biology, and we can view evolutionary function claims as parts of the functional analyses of populations.

6 GOAL-DIRECTEDNESS

I have tried to argue that part of the appeal of the SE theory comes from the fact that it appears to meet a number of desiderata for an account of functions, which are themselves shaped by the fact that modern evolutionary biology inherits an artefact model of organisms from natural theology. We must remember that this is not the only way to approach the organic world. Indeed, instead of an immediate focus on how to explain the good tools adult organisms have for jobs demanded by the environment, other biologists have instead been struck primarily by the extraordinary phenomena of organic development, maintenance and reproduction. They have been struck, that is, by the phenomena of an organism coming to be, to persist, and to reproduce over the period of time between conception and death. In many cases theorists have taken the ability of organisms to survive in spite of perturbations in their environments as definitive of life itself. Sommerhof's introductory remarks have been influential here:

> On the phenomenal level from which all science must proceed, life is nothing if not just this manifestation of apparent purposiveness and organic order in material systems. In the last analysis, the beast is not distinguishable from its dung save by the end-serving and integrating activities which unite it into an ordered, self-regulating and single whole, and impart to the individual whole that unique independence from the vicissitudes of the environment and that unique power to hold its own by making internal adjustments, which all living organisms possess in some degree. [Sommerhof, 1950, 6]

Sommerhof is here characterising whole organisms as directed towards certain goals. Parts of organisms are functional or purposive in so far as they contribute towards those goals. These goal-directed accounts appeals to what I earlier called the agent-model of functions: the appropriate way to understand organic functions is by analogy with an agent who aims at some target. The obvious goals to posit for organisms are those of attainment and then maintenance of the capacity to reproduce. Just as an agent's target can explain what the agent is doing (he is going to the shops because he is trying to get bread), so the goal-directed account

offers an additional promise not merely of characterising organic development as directed towards a target, but of explaining why organic development takes the pattern it does because it is directed at some target. So we might say that nutrients are recruited into a seedling via the roots because the seedling has the attainment of the adult form as a goal.

Today such goal-based theories are highly unfashionable ([Boorse, 2002] is an exception). Obviously the characterisation of organic development, or the maintenance of organic integrity, as directed towards a goal is almost irresistible. The question for goal-based theories of function is whether such characterisations can be justified as anything more than anthropomorphic projection, and specifically whether one can give a satisfactory account of what makes it the case that a process is directed towards some end in a way that does not fall back on something like the selected effects account. I do not propose to run through all the objections to goal-directed theories of function here, but I will give a selective review. Nissen's [1997] thorough assessment of the literature on teleology is a good source for those wishing to know more, and I have drawn on it in the following paragraphs.

Crucial to viewing development as goal-directed is the possibility of goal failure. If, for example, we want to characterise seedlings, or embryos, as directed towards the attainment of adult form, then we must be able to give an account that can fix the adult form (or the capacity for reproduction) as an organism's developmental target even when, because of disease, or some other internal or external influence, it does not reach that target. This condition immediately rules out a number of accounts of goal-directedness. We might say that organisms are directed towards whatever final states they will, as a matter of fact, assume. Goal-directedness is here accounted for merely in terms of actually realised ends. Here, however, we face a problem of how to account for goal failure. If a particular hen embryo fails to develop into a hen, shouldn't we say that it failed in its goal? But we cannot say this if we identify the goal only with how that embryo turned out as a matter of fact.

This cannot be fixed by saying that the embryo is directed towards whatever final state it is most likely to arrive at. This rules out the coherence of congenital defects, which are facts about an embryo that make it unlikely to realise the state it ought to realise, and it also forces us to say that some environmental factors that make attaining the adult state very unlikely do not perturb the organism from the adult state, but instead change the organism's goal. Perhaps all of these problems can be fixed if we say instead that an embryo is directed towards whatever outcome most other embryos of the same type actually arrive at. Yet the problem here is that in many cases we want to say that organisms are directed at ends which most of them fail to realise. Most acorns fail to turn into oaks, yet that should not rule out our identifying the production of an oak as an acorn's goal.

One might also worry that goal-directed accounts of function will suffer from problems of promiscuity. Physical and chemical systems tend to assume equilibrium states, and here, too, one need not identify the state they aim at with the state which, as a matter of fact, they will assume. A ball bearing that is rolling

towards the bottom of a cup may not reach that state simply because, let's say, Armageddon occurs before such equilibrium is attained. Now we could say that in some weak sense systems like these have equilibria as their goals. One might want to rule these out as full-blooded goals, especially if one wishes to use goal-directedness as definitional of life. The distinction that helps to mark out weak and strong goal-directedness is between systems that merely tend to some end state, and those that tend to some end state in virtue of internal compensating changes that offset the perturbing effects of environmental alterations [Sommerhof, 1950]. The ball bearing is not directed towards reaching the bottom of the cup in this sense, because the bearing is not disposed to make adjustments to its own internal structure that increase the chances of this end state being attained if the cup itself changes structure. A maturing organism, on the other hand, will make internal compensating changes in the face of perturbations to the developmental environment, which increase the chances of reproductive capacity being attained.

Thus far, we have suggested that a system directed at goal G is one such that, were environmental circumstances to change such that the attainment of some state G becomes less likely, the system would make compensating alterations in its internal structure with the result that the attainment of G becomes more likely. One should not quibble over the fact that some environmental perturbations will yield no such compensating adjustments. Obviously if one smashes a fertilised hen egg into a frying pan, nothing will happen that raises the chances of hen development occurring. The important point is merely that some more mild perturbations will yield compensating adjustments in the developing embryo. Perhaps more worrying is that we have left open in this analysis any fine specification of which end state some system is directed towards. A torpedo's trajectory is affected by a strong current, and it makes compensating changes to the position of its fins. This series of events may initially reduce, and then raise once again, not merely its chances of hitting a particular boat, but of its passing through an area of water drawn around the boat, of its producing a loud noise on impact with the boat, and of its creating a wave after impact with the boat. So all of these will count as goals on the analysis I just sketched. It is not clear to me that this is a great problem for the analysis, at least not if we are trying to demarcate systems that are goal-directed from those that are not. It does suggest, however, that if we are tempted to attribute a very specific goal to a torpedo (its goal is to destroy the boat, not to make a loud noise), we do so because we make a guess about the intention behind the torpedo's launch. If that is right, then there is no fact of the matter about whether plants are directed towards the attainment of adult reproductive states, or towards the release into the soil of a decent-sized quantity of organic fertiliser that only the death of an adult organism can yield. The compensating adjustments that occur during early development tend to make this second state likely, as well as the first.

Goal-directed accounts of function were especially popular in the middle of the twentieth century (e.g. [Rosenblueth *et al.*, 1943; von Bertalanffy, 1950; Braithwaite, 1953]), particularly in the wake of developments in the science of cybernet-

ics. This was the time when artificial systems were first constructed that could monitor their surroundings and make responsive adjustments to their behaviours: 'Some machines are intrinsically purposeful. A torpedo with a target-seeking mechanism is an example. The term servomechanisms has been coined precisely to designate machines with intrinsic purposeful behavior' [Rosenblueth *et al.*, 1943, 19].

Cybernetic accounts of goal directedness drew on the successes of these engineering endeavours, and attempted to characterise goal-directedness in terms of feedback loops. Critics were quick to pounce on problems that derived from Rosenblueth *et al.*'s assertion that 'All purposeful behavior may be considered to require negative feed-back. If a goal is to be attained, some signals from the goal are necessary at some time to direct the behavior' [Rosenblueth *et al.*, 1943, 19]. This requirement makes room for the failure of a system to attain its goal (a missile may explode before it reaches its target), but it rules out the possibility of a system having a non-existent state as a goal. A missile cannot receive signals from an object that does not exist, and an insect cannot be said to be seeking food if there is no food in its local environment. So the feedback requirement has the awkward consequence that organisms are only seeking food when food is present. Note, also, that signals of the relevant type cannot be transmitted after they are received. This rules out using a simple feedback analysis to make sense of the apparent goal-directedness of development. An adult state capable of reproduction cannot send signals to the developing embryo, because the adult state, assuming it comes to exist at all, post-dates the developing embryo.

Perhaps we could resolve this problem by looking for some internal representation, within the goal-directed system, of the goal in question. Feedback loops could then monitor whether the represented goal state is drawing nearer and direct behaviour accordingly. The main problem here, of course, is to give some sense to the idea of an internal representation of a goal state. The agent model of the organism that invites development to be characterised in terms of goal-directedness does not offer much immediate help here. On the face of it, what makes it the case that an agent has some particular goal is a plan that guides the agent's actions. And, of course, plans can be unsuccessful in their execution, or they can be directed at goals that do not exist, as when one sets out to photograph the Loch Ness Monster. Now maybe there are some animals — chimpanzees, other primates, dogs, some birds — that have intentional states, and that can form plans as a result. When we describe these animals in goal-directed terms ("The dog is trying to get the bone"), the plan-directed account of goal directedness might work, but we could not understand the unfolding of animal or plant development as guided by a plan in this intentional sense.

One possible response here is to fall back onto the selected effects account (or some other artefact-based account) as a means of making sense of developmental plans. On this view, we should think of the developmental plan for a plant as similar to the plan for a building — the plan is directed towards a particular outcome in virtue of the design process that has given rise to it. In the case

of artefacts, we specify what a plan is for by reference to the intentions of the architect; for organisms we will say that the developmental plan, itself nothing more than some arrangement of developmental materials that will help to build the organism, specifies the form of the adult organism in virtue of some aspect of its evolutionary history. For most theorists this will turn on what the elements of the developmental plan have been selected for. The adequacy of this kind of view will depend in the main on how we assess the selected effects account of functions, hence this move in effect gives up on any distinctive account of goal-directedness as the basis for function claims, relying instead on accounts of function of the form we examined in the first half of this essay.

An alternative is to think of a developmental plan as an organic analogue of an intentional state, whose properties are specified as those which are adequate to account for the repeated formation of the adult organism itself, and whose existence we should believe in simply on the grounds that it offers a good explanation of regular adult development. Paul Griffiths reminds me that Ernst Mayr may have held something like this view. Mayr defends the notion of informationally rich 'programs' that account for the goal-directedness of both individual development and individual behaviours. He acknowledges that natural selection is capable of 'producing and perfecting' these programs [Mayr, 1961, 1504], but he is resistant to making their content and existence logically dependent on their evolutionary history. Mayr's later analysis of a program as 'coded or prearranged information that controls a process (or behavior) leading it to a given end' [Mayr, 1974, 102] does not eliminate the problematic notion of goal-directedness, for the outcome that is encoded in some information-bearing structure is not merely the outcome the structure actually promotes. Perhaps, then, we should take the developmental plan or program as primitive, while offering an inference to the best explanation as the argument for its existence. It certainly is a puzzle how, in spite of a changing environment, an adult cow and bull will nonetheless be able to produce a calf that will grow up to resemble the adults. This remains a puzzle in spite of hints that discoveries in genetics have solved the problem of inheritance: no one thinks that genes are sufficient to explain trans-generational resemblance, for what effects genes have are contingent on environmental background, so that it appears that all the elements of development need to be placed appropriately for reliable development to occur. It seems we have no reason to think that chance alone would place the elements of development in the right pattern generation after generation, and nor are we aware of any general laws of nature that might guarantee that these materials will be structured so as to facilitate development. So it is legitimate to suppose that some sort of guiding plan must exist, if we are to explain the phenomena of development at all.

The main problem with this kind of account is its methodological limitations; we explain the regularity of development merely by positing some kind of plan able to bring about such development, and thereby forestall further investigation into the material causes of regular development. The account also offers no enlightenment regarding the question of goal identification. Adult development is what puzzles

us about the organic world. What allows us to identify this, rather than any other equally unlikely arrangement of organic matter, as a goal, including the arrangements of matter that constitute dead organisms, and which we tend to think of as uninteresting and unworthy of special explanation? This account posits internal developmental plans, but as yet has offered no account of what makes a plan directed towards one kind of outcome rather than another. Indeed, we might worry that it is only the fact that we find organisms puzzling that leads us to see adult states as in need of privileged explanation at all.

7 KANTIAN PROJECTIVISM

There is a line of argument that can be found in (or perhaps projected onto) Kant, which makes a virtue of these concerns. Kant suggests that we must view organisms in teleological terms — that is, as the product of something analogous to a rational plan — if we are to do biology at all. But Kant also dismisses the thought that, for humans at least, this reveals anything about the true explanation of organic form. Viewing organisms as directed towards particular goals is merely heuristic — it leads us towards the proper explanation of organic form, without containing any genuine explanation itself.[2]

Here is my best effort to use Kant's work to fashion a defensible heuristic view of goal-directedness (for a detailed philosophical treatment of Kant's work on biology I recommend [McLaughlin, 1990]). Let us begin with the problems that face a statistical account of goal failure; many young organisms do not, in fact, develop to be adults. For many species, the great majority of sperm and eggs will never combine to produce organisms at all, and many embryos will not survive beyond early developmental stages. The embryologist could, of course, attempt to give mechanical explanations for every one of these outcomes in terms of the interactions of the various causal contributors to development, and the ways in which some patterns of interaction yield short-lived entities, others yield longer-lived entities, and still others yield nothing at all. Yet such a catalogue would be unmanageable; the embryologist instead will choose to focus on some small set of outcomes as the 'normal' ones, cataloguing mechanical processes of development according to how they contribute to that normal development. Not only would the catalogue be unmanageable, the classification of some organisms' developmental trajectories as short-lived and others as long-lived itself presupposes some kind of standard for what development should look like. Development is development

[2]A note for Kantians: in this section I am taking Kant's contrast between 'constitutive' and 'regulative' principles as roughly equivalent to a contrast between principles that reflect the law-like workings of nature, and principles that reflect the methodological rules that investigators should follow to uncover nature's workings. Joan Steigerwald has warned me that this way of glossing the division between the constitutive and the regulative, as well as the consequent division it leads to between the heuristic role of teleological principles and the genuinely explanatory role of mechanical principles, may fail to do justice to the complexities of Kant's thought. I hope that readers will indulge an exploration of a loosely Kantian approach to biological teleology, even if that approach is far from Kant's.

towards some normal end state, and only by reference to this end state can we reckon some developmental trajectories truncated, and others of the proper length. By hypothesis we cannot understand 'normal' here to mean 'statistically typical'; instead, normal development means proper development. Proper development itself is then understood to mean development that proceeds as though it were guided by some successful plan.

If we think that goal-directedness is merely a projection onto nature that enables further inquiry into the mechanism of development, it is now less pressing for us to ground talk of goal-directedness by giving an account of which biological facts determine that one kind of development is proper, another improper. For the appeal to the developmental plan no longer explains development at all; rather, it tells us only which patterns of interactions between genetic and other resources we should seek to focus on as the primary basis for a subsequent attempt at mechanical explanation. It gives us a way of setting the subject matter for what will then be a causal explanation for how a fertilised egg comes to acquire new capacities over time. It would not be a threat to biology if it turns out, for example, that our perception of organic nature as uniquely goal directed (compared with, say, inorganic systems in physics) derives from a projection of our own personal striving for survival onto the organic world. That projection can still serve an essential function in providing what Kant calls a 'guiding thread' for the subsequent articulation of mechanical explanations of development. Without some way, however it is grounded, of picking on a form of organisation as the proper one, we could not pick out patterns of change as development, let alone explain them mechanically. Yet once the mechanical explanation is given the need for teleological concepts disappears, at least until we wish to draw back from a focus on the causal interactions of matter and re-characterise some collections of this matter as constituting organisms. Goal-directedness, on this view, is the ladder we kick away.

8 NATURALISM

Although Kant scholars are beginning to pay a great deal of attention to understanding Kant's views about biology, philosophers of biology very rarely defend accounts of function that take their inspiration from Kant. Part of the reason surely lies in the difficulty of Kant's work, and in the fact that it is hard to understand any part of it without already understanding all the rest. Yet I suspect another reason may lie in a perception that an account of functions should be *naturalistic*, and that naturalism entails that an account should give truth conditions for function statements that use only biological terms. This is a mistake. Roughly speaking, a naturalistic account of some phenomenon should not rely on the existence of any events or objects whose existence is denied by our best science. Projectivist accounts in ethics (e.g. [Blackburn, 1998]), for example, pass this test because although they do not identify goodness with some property endorsed by science, neither do they rely on the existence of properties denied by science in the

overall accounts they offer of ethical qualities. Similarly, the Kantian projectivist account of functions and goal-directedness which I sketched at the end of the last section does not invoke any facts denied by science in its explanation of the role and pattern of function talk in biology, but neither does it give an analysis of function talk in wholly biological terms. Of course I only sketched the account very quickly, and I certainly have not shown decisively that it makes sense of the apparently indispensable nature of heavy function talk in biology. Many difficult jobs remain for the Kantian projectivist: to show decisively why the goal-directed view of nature is not just a *useful* projection, but a *necessary* one; to explain why organisms invite the projection of goal-directedness in ways that other collections of matter do not; and to establish that we cannot give truth conditions for goal-directedness in terms of the facts about organisms that tempt us to view them as functional. Yet while these worries might cause problems for the adequacy of the account, we should not dismiss projectivist views of function on the grounds that they fail the demands of naturalism. Projectivism about biological functions is, I think, an account ripe for more detailed examination.

ACKNOWLEDGMENTS

I am grateful to Angela Breitenbach, Paul Griffiths, Mohan Matthen, Karen Neander, Denis Walsh and John Zammito for comments on an earlier draft of this paper. I owe both thanks and apologies to Joan Steigerwald: I have not been able to fully integrate her extremely detailed and illuminating comments on Kant into this revised version.

BIBLIOGRAPHY

[Allen *et al.*, 1998] C. Allen, M. Bekoff, and G.V. Lauder (eds.). *Nature's Purposes*. Cambridge MA: MIT Press, 1998.

[Blackburn, 1998] S. Blackburn. *Ruling Passions*. Oxford: Oxford University Press, 1998.

[Boorse, 1997] C. Boorse. A rebuttal on health. In Humber and Almeder (eds.), *What is Disease?* Humana Press, 1997.

[Boorse, 2002] C. Boorse. A rebuttal on functions. In A. Ariew, R. Cummins, and M. Perlman (eds.), *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press, pages 63–112, 2002.

[Braithwaite, 1953] R. Braithwaite. *Scientific Explanation*. Harper, New York, 1953.

[Buller, 1999a] D. Buller. Introduction. In D. Buller (ed.) *Function, Selection and Design*. New York: SUNY Press, 1999.

[Buller, 1999b] D. Buller. *Function, Selection and Design*. New York: SUNY Press, 1999.

[Cummins, 1975] R. Cummins. Functional analysis. *Journal of Philosophy*, 72: 741–764, 1975.

[Davies, 2000] P. S. Davies. Malfunction. *Biology and Philosophy*, 15: 19–38, 2000.

[Davies, 2001] P. S. Davies. *Norms of Nature: Naturalism and the Nature of Functions*. Cambridge, MA: MIT Press, 2001.

[Dawkins, 1986] R. Dawkins. *The Blind Watchmaker*. New York: Norton, 1986.

[Dennett, 1995] D. Dennett. *Darwin's Dangerous Idea*. New York: Norton, 1995.

[Ginsborg, 2004] H. Ginsborg. Two kinds of mechanical inexplicability in Kant and Aristotle. *Journal of the History of Philosophy*, 42: 33–65, 2004.

[Godfrey-Smith, 1993] P. Godfrey-Smith. Functions: consensus without unity. *Pacific Philosophical Quarterly*, 74: 196–208, 1993.

[Godfrey-Smith, 1994] P. Godfrey-Smith. A modern history theory of functions. *Nous* 28: 344–362, 1994.
[Griffiths, 1993] P. Griffiths. Functional analysis and proper function. *British Journal for the Philosophy of Science* 44: 409–422, 1993.
[Griffiths, manuscript] P. Griffiths. Function, homology and character individuation. Submitted to *Philosophy of Science*. Manuscript.
[Kant, 1952] I. Kant. *Critique of Judgement* (translated by James Creed Meredith). Oxford: Oxford University Press, 1952.
[Kauffman, 1995] S. A. Kauffman. *At Home in the Universe*. New York: Oxford University Press, 1995.
[Krebs and Davies, 1997] J. Krebs and N. Davies. Introduction to part two. in J. Krebs and N. Davies (eds.), *Behavioural Ecology: An Evolutionary Approach*, Fourth Edition. Oxford: Blackwell Science, 1997.
[Lewens, 2004] T. Lewens. *Organisms and Artifacts: Design in Nature and Elsewhere*. Cambridge, MA: MIT Press, 2004.
[Matthen, 1997] M. Matthen. Teleology and the product analogy. *Australasian Journal of Philosophy*, 75: 21–37, 1997.
[Matthen, 1998] M. Matthen. Biological universals and the nature of fear. *Journal of Philosophy*, 95: 105–132, 1998.
[Mayr, 1961] E. Mayr. Cause and effect in biology. *Science*, 134: 1501–1506, 1961.
[Mayr, 1974] E. Mayr. Teleological and teleonomic, a new analysis. In R. S. Cohen and M. W. Wartofsky (eds.), *Boston Studies in the Philosophy of Science Volume XIV: Methodological and Historical Essays in the Natural Sciences*. Dordrecht: Reidel, pages 91–117, 1974.
[Millikan, 1984] R. G. Millikan. *Language, Truth and Other Biological Categories*. Cambridge, MA: MIT Press, 1984.
[Millikan, 1989] R. G. Millikan. In defense of proper functions. *Philosophy of Science* 56: 288–302, 1989.
[Neander, 1991a] K. Neander. Functions as selected effects: the conceptual analyst's defense. *Philosophy of Science* 58: 168–184, 1991.
[Neander, 1991b] K. Neander. The teleological notion of 'function'. *Australasian Journal of Philosophy* 69: 454–468, 1991.
[Neander, 2002] K. Neander. Types of traits: the importance of functional homologues. In A. Ariew, R. Cummins, and M. Perlman (eds.), *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press, pages 390–415, 2002.
[Nissen, 1997] L. Nissen. *Teleological Language in the Life Sciences*. Lanham, MD: Rowman and Littlefield, 1997.
[Paley, 1802] W. Paley. *Natural Theology or Evidences for the Existence and Attributes of the Deity Collected from the Appearances of Nature*. London: Faulkner, 1802.
[Rosenblueth *et al.*, 1943] A. Rosenblueth, N. Wiener, and J. Bigelow. Behavior, purpose and teleology. *Philosophy of Science*, 10: 18–24, 1943.
[Sober, 1984] E. Sober. *The Nature of Selection*. Cambridge, MA: MIT Press, 1984.
[Sommerhof, 1950] G. Sommerhof. *Analytical Biology*. Oxford: Oxford University Press, 1950.
[von Bertalanffy, 1950] L. von Bertalanffy. An outline of general system theory. *The British Journal for the Philosophy of Science*, 1: 134–165, 1950.
[Waddington, 1957] C. H. Waddington. *The Strategy of the Genes*. London: George Allen and Unwin, 1957.
[Williams, 1966] G. Williams. *Adaptation and Natural Selection*. Princeton: Princeton University Press, 1966.
[Williams, 1992] G. Williams. *Natural Selection: Domains, Levels and Challenges*. New York: Oxford University Press, 1992.
[Wright, 1973] L. Wright. Functions. *Philosophical Review* 82: 139–168, 1973.
[Wright, 1976] L. Wright. *Teleological Explanation*. Berkeley: University of California Press, 1976.

This page intentionally left blank

BIOLOGICAL APPROACHES TO MENTAL REPRESENTATION

Karen Neander

1 THE PROBLEM OF INTENTIONALITY

Teleosemantics is a biologically inspired, some of us would like to say biologically informed, approach to the philosophical problem of intentionality. Whether it will prove right or fruitful in the long run is an issue with significant implications not only for issues in philosophy of mind but also for the relations between biology and psychology and for the nature of neuroscience.

While recognition of the problem of intentionality is arguably ancient, it has most famously been stated by the 19th century Austrian philosopher, Franz Brentano. As Brentano [1874, p. 112] says,

> Nothing can be judged, desired, hoped or feared, unless one has a presentation of that thing. ... [W]henever we desire something ... we have before our minds that which we desire.

In Brentano's terms, desiring wine involves a presentation of wine. More generally, he says, you cannot have any mental attitude toward a thing unless it is present to your mind. You cannot hope, doubt, taste, believe or remember that the wine is fruity unless the wine and its being fruity is present to your mind. However, as Brentano emphasized, a thing need not be present in the ordinary sense to be present to your mind. The wine must be present if you are to taste it, but you can hope, doubt, believe or remember that it is fruity in its absence. We can think about Black Holes and polar bears when none are nearby. We can wish for the fountain of youth or a cure for cancer even if there is no fountain of youth or cure for cancer. In what way, then, is the object of our thought 'present' at all?

By a presentation of a thing Brentano tells us that he means reference to content. The thing is 'present', as we would nowadays express it, only in so far as it is represented. But these are just more words to describe the same phenomenon: How do minds and brains represent things? What is reference to content?

Teleosemantic theories (extensively defended by Millikan [1984; 1989a; 2000], Papineau [1984; 1987; 1993], Dretske [1986; 1988; 1994], and others) are a diverse range of answers to this question, but each teleosemantic theory maintains that the answer involves a (in a certain sense to be explained) normative and teleological

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

notion of function, or in other words a notion of function that underwrites talk of malfunction and that is construed as a notion of what traits or items are for. This is usually understood in terms of an etiological theory, often referred to as a selected effect (SE) theory of functions [Wright, 1973; Millikan, 1989b; Neander, 1991] according which functions are what items were selected for. In the case of artifacts, the functions are said to result from intentional selection (e.g., what something was designed for). In the case of natural functions, the functions are said to result from a natural process of selection (e.g., ordinary natural or genic selection). In his entry on functions in this volume, Timothy Lewens, calls this kind of function talk "heavy function talk".

There is no knockdown argument for any theory of mental content and if a version of teleosemantics is to win its contest with competing theories it has to win on points, all things considered. However, there is a methodological reason for thinking that some version of teleosemantics is probably correct, or more modestly that teleosemantics at least deserves an especially long, hard look. I explain this reason here, after first saying more about the problem of intentionality and the teleosemantic response to it, and then I review some problems that teleosemantic theories face.

2 FUNCTIONAL AND SEMANTIC NORMS

There are a number of different teleological theories of mental content. I don't explain any of them here in full detail, but it's central to all of them that a certain normative notion of function underwrites a certain normative notion of content. I'll say something about the latter notion, something about the former notion, and then something about how the two can be brought together.

Content is said to be normative because some mental states that have content can represent correctly or they can misrepresent. The capacity to misrepresent distinguishes representing from mere carrying of information. One seminal presentation of this idea is found in H. P. Grice [1957]. As he says, dark clouds can "mean" that a storm is on its way, but only if a storm is on its way. While a hiker can misunderstand the significance of the clouds, the clouds themselves cannot misrepresent an approaching storm. If there was no storm approaching, the clouds never meant one was. Similarly, Johnny's spots can "mean" that Johnny has the measles, but only if Johnny has the measles. The doctor can misunderstand the significance of the spots, but the spots themselves cannot misrepresent the measles. If Johnny doesn't have the measles, his spots never meant he had.

Grice contrasts the significance of natural signs (which he calls "natural meaning") with the meaning of the doctor's words (which he calls "non-natural meaning"). The doctor's words, "Johnny has the measles", can mean that Johnny has the measles even if he does not have them. While Grice does not focus on the fact, the same is true of the doctor's thought that Johnny has the measles. The doctor's thought can have the content that Johnny has the measles even if Johnny doesn't. In sum, linguistic utterances and mental representations do more

than carry information in the way that natural signs do — they represent. And representation, unlike the mere carrying of information, is normative. Nothing can carry the information that P unless P, but representations can, in general, represent that P even if not-P [Dretske, 1986].

Despite Grice's reference to the second kind of meaning as "non-natural meaning" not all of it can be conventional, or dependent on our intentions, although this might well be the case, as a Gricean analysis claims it is, for linguistic meaning. Ultimately, some of our intentional mental states must get their meaning without having to derive it from the meaning of other intentional mental states (on pain of either circularity or infinite regress). Teleological theories of mental content, like other attempts to provide what are referred to as naturalistic theories of mental content, try to explain how this is so.

Cognitive systems are systems adapted for producing and processing internal states that carry information, and for using these states to adapt the bodies in which they are situated to the environments in which they in turn are situated, and vice versa. But, as we have just seen, minds and brains do not merely carry information about their environment. They represent their environment. And this introduces semantic norms (of correctness and incorrectness) of which some account is needed. Teleosemantic theories propose that psycho-semantic norms ("psycho" for psychological) derive from functional norms that in turn derive from the selection processes that were (in part) responsible for the development, phylogenetic or ontogenetic, of these cognitive systems.

The relevant notion of content is said to be normative because a state with such content can, in general, represent correctly or incorrectly. The relevant notion of function is said to be normative because a trait with such a function can, in general, function properly or malfunction. A robin's wing has the function of flight even if it is broken and cannot fly. A heart has the function of pumping blood even if it is suffering cardiac arrest and cannot pump. According to an SE theory of functions, a trait is said to have a natural function, x, if traits of that type were selected for x-ing by a natural (here, meaning non-intentional) process of selection. To a first approximation, according to an SE theory, a trait is functioning properly if it has the capacity to do that for which traits of the type were selected, and it malfunctions if it lacks that capacity.[1]

Some dislike talk of functional norms because they like to preserve the word "norm" for prescriptive contexts. However, its use is not in general restricted in this way, since we sometimes talk of statistical norms and these are clearly not prescriptive per se (it is not necessarily ideal to be of average height, looks or intelligence). In any case, the terms are unimportant. What matters is that those

[1] As I'll have reason to note again in the last section, there is more than one correct way to describe a trait's function because traits are selected for a concertina of effects that ultimately result in gene (meme, etc.) replication. A trait does not malfunction merely because it cannot bring about this full concertina of effects. A woman's ovaries do not malfunction just because her fallopian tubes are blocked and so she cannot conceive. They malfunction only if they lack the capacity to contribute what they were selected for more particularly contributing to an enterprise that involves a great many other co-adapted parts [Neander, 1995].

who support teleosemantics are not claiming that functional norms (or "quasi-norms" if you wish) are prescriptive; they are claiming that they are descriptive. They are claiming that they can be analyzed in terms of what there was selection for, not by God or by humans, but by natural processes of selection. If the norms of content were prescriptive, and functional norms were not, then that might be a problem for teleosemantics. Or if functional norms were prescriptive, and the norms of content were not, then that might also be a problem. However, according to teleosemantics, both are descriptive. On this view, neither claims about normal functioning nor claims about correct representation necessarily imply claims about what should be the case, either morally or pragmatically speaking. And while it is true that, when we speak of functions, we often find ourselves speaking of what traits are "supposed to do", or even of what they "ought to do", this is just sloppy or colloquial or metaphorical talk of what they were selected for doing, on an SE account.

We can get some sense of the role that the notion of an SE- function is intended to play in a theory of content if we start with a crude causal theory of mental content, and see how the problem of accounting for the possibility of misrepresentation arises for it. The crude causal theory says that a representation represents whatever causes it. It says that a token representation of type $\underline{R}$ correctly represents Cs iff (if and only if) Cs and only Cs cause $\underline{R}$s. For example, it says that CAT-representations represent cats iff it is the case that cats and only cats cause CATs.

This theory has many well-known problems, one of which is the problem of error.[2] Consider an occasion on which, when walking down the street, I see some crumpled newspaper blown by the wind as a cat slinking. Since the crude causal theory says that the content of a representation is whatever causes it, and the newspaper is now among the causes of my CAT-representations, it entails that the content of my CAT-representations includes the newspaper. This is part of what is (after [Fodor, 1990a]) referred to as the "disjunction problem". The crude causal theory entails that my CAT-representations refer, not just to cats, but instead to cats *or* newspapers *or* anything else I happen to mistake for a cat. Of course, this is intended as a reductio of the theory since it shows that the theory makes error impossible.

There are a number of directions one can take here. One might decide that any attempt to develop a causal-information based psychosemantics is doomed to fail. That would be too hasty, but this is a direction one might decide to take anyway, once further considerations come into play. Another approach is to try to distinguish the content-conferring causes from the non-content-conferring ones. A teleological theory of mental content might take either tack, but I follow the

[2] An alternative to the idea that mental representation starts with information is an isomorphism theory. And an alternative to the idea that information is based on causation is that a notion of indication suffices. There are teleosemantic theories that use both of these ideas. The problem of error could have been introduced using these ideas too, but for simplicity I here stay with a causal-informational approach.

second here for illustrative purposes.

If we think of functions as selected dispositions, we can suggest that the content of a representation, $\underline{R}$, is not what causes $\underline{R}$s so much as what $\underline{R}$s have the function of being caused by. The idea is better expressed in terms of the functions of the mechanisms that produce the representations. For instance, we might suggest that $\underline{R}$-representations refer to Cs if certain mechanisms were adapted for responding to Cs by producing $\underline{R}$s, or in other words if they were (in part) selected for their disposition to be caused by Cs to produce $\underline{R}$s. My perceptual mechanisms plausibly have the function of producing CATs in response to cats but they do not plausibly have the function of producing CATs in response to newspapers. In this way we have a way of explaining why CATs refer to cats and not newspapers, despite the fact that both of them can cause CATs.

This is at best a second step (and it might be a misstep) toward a more complete theory. Many problems remain. In the last section I outline a few of them, along with some possible responses. However, this simple teleosemantic theory illustrates an important insight, which is that by employing a descriptive notion of a natural function, we can start to provide a distinction – a descriptive and scientifically respectable distinction — between what happens in representational systems and what is, so to speak, "supposed" to happen — we can start to develop the beginnings of a naturalistic analysis, in other words, of correct and incorrect representation.

3 THE GREAT DIVIDE

The theory described in the previous section is a simple one, with many problems. But before we look at some of these, I want to suggest that we have a methodological reason to take an especially close look at teleosemantics. The place to look for this reason is neuroscience, which is where talk of representation first enters the scientific picture if we start with sub-atomic physics and travel toward psychology. However, this claim contradicts the popular idea that there is a Great Divide between the "intentional" and "non-intentional" sciences, and so let's take a moment to consider this idea.

By "intentional sciences" I refer to sciences that study intentional systems as intentional systems. Psychology and the social sciences are examples. Psychologists study the dynamics of intentional mental states within people and within small groups of people and the social sciences, such as history, politics and economics, study the dynamics of populations of people, construed as agents. By "non-intentional sciences" I refer to sciences that either do not study intentional systems at all (e.g., geology and cosmology) or do not study them as intentional systems (e.g., physics and organic chemistry). In the latter kind of science, an intentional process might be described, but it is not described as an intentional process insofar as its intentional content plays no role in the description. Thus a brain process might be described in terms of ion channels and synaptic excitation and inhibition without mention of the fact that it is, say, a perception of a duck

or a memory of last night's dinner.

Just as the notion of function is a peculiarity of biology relative to the more basic natural sciences, the notion of a representation seems to be a peculiarity of psychology (and the social sciences) relative to biology and the more basic natural sciences.

On the Cartesian picture, the Great Divide between the intentional and non-intentional sciences is thought to have an ontological basis, because on that picture bodies and brains are physical systems and minds are not, and so little continuity is to be expected between the natural sciences, which aim at explaining the former, and psychology and the social sciences, which aim at explaining the latter. With the demise of Cartesian Dualism, the case for integrating psychology with the natural sciences seemed to improve. According to Ontological Behaviorism, mental terms have a behavioral analysis. And according to Reductionist (type type) Physicalism, psychological properties are neurophysiological properties, and bridging the divide is merely a matter of discovering these identities. However, neither behaviorism nor reductionist physicalism remained very popular for long.

With the arrival of the view that is commonly (but in this context a little misleadingly) known as Functionalism, the Great Divide reappeared, as if out of a mist. Functionalism consists of two main theses. The first is that mental states cannot be reduced to brain states because mental states are multiply realizable and the second is that mental states can be characterized in terms of their functional role. Pain in humans might be realized by a neurophysiological state such as (toy example) C-fibers firing. But, says the functionalist, this does not preclude other kinds of creatures that lack C-fibers from experiencing pain: Pain in octopi or pain in Martians might be realized by different kinds of physical processes. What pains have in common, on this view, is their characteristic disposition: Pain is the kind of state that is characteristically caused by bodily damage and that characteristically causes certain other specifiable inner states and motor outputs.

What "characteristically" means is not usually made explicit, but the kind of function talk that is used in Functionalism is not usually understood as talk of normative and teleological functions. Such talk is usually understood as talk of what are sometimes known as CR functions, about which more shortly.

Functionalism, as it is standardly understood, is consistent with physicalism, but it is also consistent with the Great Divide, especially since intentional mental states are expected to be even more diverse in their realizations than phenomenal states (like pain) are. Not only might a desire for peace be realized by different states in humans and in aliens from Alpha Centauri, but it might also be realized by different states in different humans, or by different states in the same human, as a child develops into an adult, or perhaps even on different occasions, from one day to the next. And just as we can study computer programming and data systems without studying hardware engineering, we can study human psychology without studying human neuroscience, or so the reasoning goes. In fact, it is argued that we must preserve the autonomy of cognitive science, because it is only if we abstract away from the diverse details of neurological realizations that we

can see the important intentional generalizations.

On this view, neuroscience is a non-intentional science. We often see this view expressed in philosophy of mind. Consider, for example, Stephan Schiffer's puzzle over Ava, who is about to cross the street when she sees a car coming and steps back to the curb. As Schiffer remarks, there was within Ava an unbroken chain of neural events, beginning with the stimulation of receptor cells in her eyes by light reflected from the oncoming car and ending with her stepping back. Each of these neural events was a cause of the one following it and so of her stepping back. This, Schiffer says, conforms to certain laws:

> As these laws are laws of neurophysiology, laws that pertain to the electrochemical properties of the neural events they subsume, this premise implies that there is a sufficient explanation of Ava's stepping back wholly within the language of physical science. [Schiffer, 1987, 146–7]

What need is there, Schiffer goes on to ask, for an explanation in intentional terms? What explanatory role can representational properties play? Note that Schiffer is not questioning the need for neurophysiological explanations in the face of the in-principle sufficiency of more basic physical explanations, he is questioning the need for intentional explanations, in the face of the sufficiency of neurophysiological explanations, on the assumption that neurophysiological explanations are not intentional.

The same assumption crops up in early arguments for eliminative materialism by Paul Churchland. Churchland [1981], for example, argues that folk psychology is a moribund research program and suggests that it might ultimately be replaced by mature neuroscience. Folk psychology is the psychology we learn in the schoolyard rather than in the schoolroom; it is our ordinary everyday understanding of minds, commonsensical rather than scientific. Churchland argues that there is much about the mind on which folk psychology is silent. It cannot explain mental illness, the function of dreams, pre-linguistic learning, or the formation of a 3D visual image from a 2D retinal array.

He argues that it is, in the terms of Imre Lakatos, a barren and stagnant research project, one that has made no progress for thousands of years. Folk psychological terms, such as "belief", "desire", "judgment" and "knowledge" are, he maintains, defined by their role in the predictive and explanatory laws of folk psychology — the laws, such as they are, that are used "in the market place" to explain and predict human behavior. Since this body of laws could prove to be radically wrong, the categories defined in terms of them could turn out to have no home in mature scientific psychology.

This is considered a central argument for eliminative materialism, and more specifically for the claim that talk of intentional mental states might be eliminated from science. However, this is only an argument for eliminative materialism on the assumption that neuroscience itself is non-intentional. If it were intentional, then the ascendance of neuroscience would be a vindication of intentional realism.

4 BRIDGING THE DIVIDE: NEUROSCIENCE

This view of neuroscience needs to be discarded. Neuroscience is an intentional science. It counts as an "intentional science" if by this we mean a science that studies representational states as representational states (and do not, say, equate intentional states with folk psychological states).

It's probably right to say that neuroscience began with a bottom-up approach, first focusing on molecular and cellular phenomena, and only later considering the representational functions of neural components. However, this focus was never exclusive, and as neuroscience matures and becomes more thoroughly integrated with the rest of cognitive science, it is replete with talk of the representational functions of neural components.[3] As they are practiced today, there is no sharp distinction to be drawn between neuroscience and cognitive science more generally. It is a difference in emphasis only. Neuroscience seeks to describe and explain our neurological processes and understand how they relate to our cognitive capacities, whereas cognitive science seeks to describe and explain our cognitive capacities and how they relate to our neurological processes.

To understand why neuroscience is representational we need to consider the nature of physiology in general and neurophysiology in particular. In brief, physiology is the study of the functional organization of living systems. So we cannot complete our study of brain physiology until we have described its functional organization. Since neural components have representational functions, it follows that we cannot complete neurophysiology until we have described the representational functions or the information-processing functions of neural components.

But how does this amount to motivation for teleosemantics? The answer is simple. Central to physiology in general and to neurophysiology in particular is a notion of function that is normative. If the SE analysis of it is the best analysis of it, neurosphysiology is already steeped in talk of SE-functions, and indeed of information-processing SE-functions. This is not the place for a general defense of an SE analysis of normative functions. Such defenses have been given elsewhere (see [Neander, 1991], and [Neander, 2005] for a response to [Lewens, 2004]). However, I need to say something about the role of SE-functions in physiology in general and in neurophysiology in particular, since a proper understanding of this role is needed to see why we should take teleosemantics seriously.

It has become a near commonplace in philosophy of biology to think that there is a division of explanatory labor between SE-functions, on the one hand, and CR-functions, or causal-role functions, on the other. SE-functions, as I said in the opening section, are what traits of a type were selected for. CR- functions, in contrast, are dispositions, or causal roles that items are disposed to play. SE-functions are normative in the sense that an item can have an SE-function that

[3]Now, by the way, the Churchlands would agree with this. Paul Churchand [1995] offers a similarity-based vector analysis of how neural nets represent, and Pat Churchland [2002, 273–318] explains why, in her view, neuroscience needs a notion of representation and why, at least so far, no viable alternative is in sight.

it lacks the disposition to perform. CR-functions are not normative, because an item cannot have a CR-function that it lacks the disposition to perform, and so CR-functions do not underwrite talk of malfunctioning.

It has, as I say, become a near commonplace to think that there is a division of explanatory labor between SE-functions and CR-functions, and the usual understanding of this division seems sensible on the surface. The usual understanding is that SE- functions explain how items with functions evolved, since SE- function ascriptions tell us what items were selected for (i.e., that which, after selection, becomes their function), whereas CR- functions explain how complex organic systems operate, since they tell us what items are disposed to contribute to the operation of the systems in which they are embedded. On this view, SE- functions play an important role in evolutionary explanations, whereas CR-functions play an important role in operational explanations in physiological biology.

This seems sensible on the surface but there are two main problems with it. One is that, although it is true that SE- function ascriptions make evolutionary claims, this does not amount to an important explanatory role for them. On this view, all the SE-function ascriptions are doing is summarizing evolutionary claims that are reached independently, and such claims could easily be made in other terms (instead of saying that t has x as its function, we could say that t was selected for x-ing). On this view, the notion of an SE-function is one that is easily eliminated, so if SE-functions play an important explanatory role in evolutionary biology, this is not it.

The second problem is that it is physiological biology, not evolutionary biology, that most obviously makes use of a normative notion of function. Physiology is the study of the functional organization of living systems and physiology is steeped in talk of organic systems' functioning properly or malfunctioning, of organic systems' functioning normally or abnormally, of functional impairment, dysfunction, and so on and so forth. Anyone who doubts this should just take a look, since it can be seen at a glance at any physiology, or for that matter neurophysiology, text. This normative notion is not the notion of a CR-function. It is, according to many, the notion of an SE- function. Therefore, there must be something wrong with this standard account of the division of labor. Unless biologists are terribly confused, normative functions belong in operational explanations of how organic systems work, or fail to work.

What are they doing there? The explanatory role of normative functions in physiological biology is in part the role attributed to CR-functions by Robert Cummins, but it adds an element of idealization. This is how Cummins puts it:

> The biologically significant capacities of an entire organism are explained by analyzing the organism into a number of "systems" — the circulatory system, the digestive system, the nervous system, etc., — each of which has its characteristic capacities. These capacities are in turn analyzed into capacities of component organs and structures. Ideally, this strategy is pressed until the analyzing capacities are amenable

> to the subsumption strategy. We can easily imagine biologists expressing their analyses in a form analogous to the schematic diagrams of electrical engineering, with special symbols for pumps, pipes, filters, and so on. [Cummins, 1975, 760–1]

This is all true but it omits the peculiar kind of idealization that is used in biology in providing such functional analyses. When biologists provide such explanations, they make sure to distinguish between normal and abnormal capacities, or in other words they make sure to distinguish between what different parts of a system contribute to the overall operation of the system as a whole when they are functioning normally or properly and what they contribute when they are not functioning normally or properly.

This point is not new. It was familiar to philosophers of biology before Cummins wrote his piece, but it has often gotten lost since Cummins wrote his piece, and it is important to the present discussion. To be clear, on its own it does not establish the need for an SE-analysis. However, it is worth noting that SE-functions mesh well with this idealization task. Organic systems are tremendously complicated organized systems and they could not exist except for the fact that they are the products of selection. Furthermore, they mostly only operate well — in terms of survival and reproduction, and in terms of the vast array of capacities that contribute to them — to the extent to which each of their intricately co-adapted parts is performing the causal role for which it was selected. That's because of how selection works: basically (I simplify, but not in ways that need concern us here) it tends to keep the parts that contribute to survival and reproduction and it tends to eliminate the parts that hinder them. A further consideration is that there are many, often billions of instances, of each type of system that the physiologist seeks to describe: consider all the humans that have ever lived (96,100,000,000, according to one estimate that I have read) let alone all the ants or fruit flies. At least some of the immense variation among individuals can be set apart in a theoretically significant fashion, when describing a type of system (a "species design", so to speak) by distinguishing between normal and abnormal variation.

Neurophysiology is no different. It also aims to provide an idealized functional analysis of a tremendously complicated organized system. Only in this case, as we currently understand it, it is a cognitive system designed for processing and using information. Therefore, a normative notion of a function, of the kind used in ordinary somatic physiology, is already deeply implicated in cognitive science. It is already intimately involved in scientific talk of information-processing functions. On the assumption that the normative notion of a function is the SE-notion, the general teleosemantics enterprise is a parsimonious one. It makes good sense to see how far we can go on the basis of the notions already in play.

5 SOME PROBLEMS FOR TELEOSEMANTICS

If the argument just offered is along the right lines, there is reason to persist in trying to produce a good teleosemantic theory. In this last section I outline four problems for teleosemantics, along with sketches of possible responses to them. My treatment is sketchy and partial, of necessity. But my aim is merely to provide readers with some idea of the challenges that teleosemantics theories face and an idea of a few of the resources that they have available to them.[4]

(1) Novel concepts. It is extremely unlikely that all of our concepts are innate. For example, it is extremely unlikely that even the folk species concepts (CAT, SHEEP, WOMBAT, etc.) are innate, let alone that more obviously modern concepts, like ELECTRON and JET-PLANE, are.[5] What it means to say that a trait is innate is a delicate issue, but it will suffice for present purposes to say that a concept is innate if there was selection specifically for it, and if the selection was ordinary natural (genic) selection. Acquired as opposed to innate concepts are a prima facie problem for teleosemantics if it relies on a notion of function that in turn relies on the idea that functions are what items were selected for by ordinary natural (genic) selection.

How do proponents of teleosemantics respond to this? One type of response is to appeal to other forms of natural (here, meaning non-intentional) selection to underwrite the relevant notion of a proper function. Meme selection [Millikan, 1984] and conditioning [Dretske, 1986] are sometimes mentioned in this context, and neural selection might also be a possibility. A danger with this approach is that the notion of selection becomes so vague that it threatens to become vacuous, but with work it might be shown that this danger can be met.

A more complex but, to my way of thinking, equally plausible idea is that the content of acquired concepts is not derived straightforwardly from a single selection process, but that selection is involved nonetheless. For example, it is more plausible that a general disposition to form animal species concepts is innate than that all of our particular animal species concepts (CAT, SHEEP, WOMBAT, etc.) are. This disposition might take the form of a disposition to store memories (schematic or detailed) of animals of the same species within a specialized type of "mental folder" (a metaphor for a functional component). Along the same lines, we might have an innate disposition to form other kinds of concepts too, such as concepts of artifacts, natural kinds, and individuals, so that we come equipped with a disposition to form different kinds of concepts for different kinds of kinds

[4] I omit in what follows a resource that some proponents of teleosemantics think terribly important: the role of the representation's consumer. I have never been persuaded that this does any important work for teleosemantics that cannot be done without it, but see Millikan [1989a] and Matthen [2006] for the opposed view.

[5] There is much debate on this, but this claim seems plausible. After all, infinitely many different species are possible. And assuming that we could see each of them, and interact with each of them in the usual way, there is every reason to think that we could acquire a concept of whichever species happen to be present in our environment. Since we cannot possess an infinite number of concepts innately, we must therefore be able to acquire particular species concepts ontogenetically.

of things. Genic selection might then partly determine the content of a particular animal species concept. It might determine that the content of (e.g.) CAT is some species or other, since that is the specialized type of "mental folder" that is used. Then the question becomes how the content of the concept is tied to the particular species involved.

At this point a teleological theory of content could be developed in a variety of ways. It might appeal to an alternative form of selection, as mentioned before. Perhaps conditioning or neural selection tunes the specialized "mental folder" to the particular species. Or it might appeal instead to a further causal condition. For instance, the idea might be that each new animal-species "mental folder" refers to the species to which its triggering specimen belongs.

(2) Non-existent intentional objects. However, some concepts cannot be so readily handled. For example, no matter what kinds of actual selection processes we allow, our perceptual mechanisms have not been selected for producing a UNICORN in response to a unicorn, and nor was any actual unicorn a triggering cause of our UNICORN concept. Since there are no unicorns, neither we, nor our ancestors, could have causally interacted with them. A parallel point applies to all concepts of things with which we have not or cannot have causally interacted at the time of our reference to them: e.g., TOMORROW.

One response to this, insofar as it concerns non-existent intentional objects, is to deny that concepts of them have referential content of the kind analyzed by teleosemantics. Those who do this can still allow that UNICORN has something like a Fregean sense — a mode of presentation, or a cognitive content. However, this is unsatisfying in part because it doesn't generalize (e.g., to TOMORROW).

Another option is to concede that UNICORN and TOMORRROW et al cannot be simple. To say that a concept is not simple is not the same as saying that it is not innate. A concept can be simple without being innate (I described two ways in which this might happen in the preceding sub-section). If a concept is not simple, it derives its content from the contents of other concepts, and this is the possibility I am now pointing to. The contents of UNICORN and TOMORROW on this view would derive from the contents of other concepts (UNICORN from HORN, MIDDLE, FOREHEAD, etc., and TOMORROW from DAY, AFTER, TODAY, etc.).

Examples like UNICORN and TOMORROW are not devastating for teleosemantics unless there is good reason for thinking that they must be simple. Arguments for conceptual atomism [Fodor and Lepore, 1992] could be seen as constituting such reasons but they are not strong enough for this. This is a large claim that cannot be adequately defended here, but bear in mind that the arguments for conceptual atomism are negative arguments against particular theories of complex concepts. Consider, for example, the view that complex concepts have definitions consisting of necessary and sufficient conditions for the application of the concept. A major problem with this view is that few of our concepts seem to have good definitions. But the fact that few concepts seem to have good definitions is no reason to think that none do. And TOMORROW would appear to be an excep-

tion: Tomorrow is the day after today.[6] And so, as long as DAY, TODAY and AFTER have instantiations with which we have interacted, it is not clear that TOMORROW presents a deep problem.

Perhaps definitionism does not work for all complex concepts. Perhaps, it does not work for UNICORN. But the philosophical and psychological literature on concepts offers other ideas about how complex concepts could be structured and it could be that different complex concepts have different kinds of structures.[7]

(3) Functional indeterminacy [Dretske, 1986; Fodor, 1990a]. A further problem arises for teleosemantics because there are different ways of describing the dispositions for which traits have been selected. One reason for this is that traits are selected for a concertina of effects. They are not as a rule selected for simply doing x, but for doing x^1, which brings about x^2, which brings about x^3, which ultimately tends to contribute to, say, gene replication. Another reason is that there are can be different but extensionally equivalent descriptions available for the environmental features with which a cognitive system interacts. The much-discussed case of the frog's prey-capturing mechanism has been used to illustrate both parts of this problem.

The frog's perceptual system is adapted for responding to stimuli that are small, dark, and moving past its retinas. In a normal frog, such stimuli produce a characteristic neural event in its optic tectum (here, $\underline{R}$) which in turn causes the frog to turn to face its prey and try to catch it by approaching it, and snapping its tongue or jaws. This in turn allows the frog to ingest food, digest it, and circulate nutrients, which then promote the full range of bodily functions. Thus the frog's perceptual mechanism was selected for a complex causal role — for its disposition to respond to stimuli that are small, dark and moving, but also for its disposition to guide the frog's snapping, help it to catch food, provide nutrients to cells, and so on. On the simplifying assumption that, in the environment in which frogs evolved, small, dark, moving dots were co-extensive with flies and frog food, we can also see how different descriptions of co-extensive environmental features can

[6]The fact that this definition does not unequivocally settle all possible questions of application is not a reason to reject it. A good definition need only settle issues that are genuinely settled by the concept being defined.

[7]Perhaps teleosemantics would be in more serious trouble if it could be shown that some representational simples have non existent intentional objects. Georges Rey [manuscript] argues that this could be the case. He cites psychological theories (like those of [Marr, 1982], and [Biederman, 1990]) that propose that perceptual processing employs "geons", or volumetric shape representations, e.g., of cones and cylinders. Suppose for example, that CIRCLE was a simple of our visual system. Since there are no actual perfect circles, CIRCLE is an empty concept. It is debatable whether we possess a visual simple of a circle that is as exacting as Rey's argument requires. But extant versions of teleosemantics imply, not only that do we not have such a simple, but that we cannot have one.

I'm not sure how troubled the proponents of teleosemantics should be about this possibility. If there really were a need to admit visual simples of Platonic ideals, perhaps accommodation could be made. We would need to examine the details of the case, but neural components can, for instance, be adapted for responding more or less enthusiastically to more or less good instances of certain properties. Maybe there is room for actual selection history with actual non-ideal instances to ground reference to ideal ones or approximations of ideal ones.

come into play. In that case we can say that, while the frog's perceptual system was adapted for producing $\underline{R}$s in the presence of small, dark, moving things, it was also adapted for producing $\underline{R}$s in the presence of flies or frog food. So, it has been argued, an appeal to the function of the frogs' detection mechanism does not give us determinate content for $\underline{R}$s. Depending on how we describe its function, we get different content ascriptions. Described one way, the content is small, dark, moving things, and described another way it is flies or frog food or something nutritious.

Again there are different ways of tackling this problem, and how one goes about it depends in part on what one thinks the correct content is in this case, and in similar cases. The main point to note, however, is that it is wrong to assume that, if functions are indeterminate in this way, then content derived from teleological theories must be indeterminate in this way too. This doesn't follow, simply because teleological theories can do more than mention functions. The content one derives from a teleological theory depends on how functions are (or selection is) put to use in the theory.

Millikan's [1991] position on this is the one that is probably best known. She claims that the content is frog food. Her teleosemantic theory is quite complex, but the gist of it, regarding this case, is that the content of the frog's $\underline{R}$s is frog food because it was $\underline{R}$'s mapping on to frog food, as such, rather than $\underline{R}$'s mapping on to small, dark and moving things, as such, that explains why there was selection for the use of $\underline{R}$s. Talk of explanation can probably be eliminated here, because what is at issue is what was essential for a contribution to the proliferation of the frog's genes. Millikan seems to be thinking something like this. Had the frog swallowed something that was frog food, the fitness value of its $\underline{R}$-induced snapping would have been the same, regardless of whether it was small, dark and moving. Whereas had the frog swallowed something that was small, dark and moving, the fitness value of it's $\underline{R}$-induced snapping would not have been the same, regardless of whether it was frog food.

Others argue that the correct content is the sign stimuli: small, dark, moving things. Neander [forthcoming] argues that this content is better suited to mainstream (information processing) neuroethological explanations of the frog's perceptual capacities, and Paul Pietrosky [1992] argues for analogous content in an analogous case for more intuitive reasons. Some of Fred Dretske's writings also support this result (e.g., see [Dretske, 1994]) although to see this we need to note that frog food did not strictly co-vary with small, dark, moving things in the environment in which frogs evolved: the flies sometimes stood still; bits of bark or small shadows sometimes moved. In these places Dretske speaks of the content of $\underline{R}$s as what $\underline{R}$s were recruited for indicating (or sometimes, he says, "maximally indicating") where $\underline{R}$s indicate Cs iff if there is an $\underline{R}$ then there is a C in the relevant environment. (This manner of addressing the problem speaks only to the first part of it, not to the second.)

The simple teleological theory with which we began also gives the sign stimuli as the content, with only minor tweaking. Remember that it says that $\underline{R}$-

representations refer to Cs if mechanisms were adapted for being caused by Cs to produce $\underline{R}$s. We get the desired result if we instead suggest that $\underline{R}$-representations refer to Cs if mechanisms were adapted for being caused by Cs qua Cs (i.e., in virtue of their being Cs) to produce $\underline{R}$s. That is, the frog's $\underline{R}$s refer to small, dark, moving things, on this proposal, if the frog's perceptual mechanism was selected for being causally sensitive to the property of something's being small, dark and moving.

(5) Swampman. In his original manifestation, Swampman is a doppelganger (molecule for molecule duplicate) of Donald Davidson at a time ([Davidson, 1987], though earlier versions of Swampcreatures exist, e.g. [Boorse, 1976]). Where Swampmman differs from Davidson is in his history. Swampman comes about by sheer accident as a result of a random collision or coalescing of sub atomic particles and his resemblance to Davidson is sheer coincidence. According to a teleosemantic theory, Swampman has no intentional content, since he has no history of the appropriate sort: no evolutionary history of selection whatsoever, ontogenetic or phylogenetic. The problem is that it is intuitive to think that Swampman could think about things. Or it is at least intuitive to think that it will seem to Swampman that he does. Does he not see the world about him? Does he not think about it? Indeed, how could it even seem to him that he has such mental states unless he at least has *that* intentional state?

One issue is how seriously we should take this type of imaginative counter-example. Arguably, the objection fails for versions of teleosemantics that are offered as attempts to describe the a posteriori necessary conditions for intentional kinds, which are psychological kinds that are individuated by their contents (e.g., perceptions of red, and a child's concept of an object). Teleosemantics treats intentional kinds as analogous with biological kinds. Just as it is not a strong argument against an historical definition of a species concept that it precludes Swampman from being a member of Homo sapiens, it does not seem to be a strong argument against an historical theory of mental content that it precludes Swampman from having content, if these are similar kinds of kinds (see [Dretske, 1996]; [Neander, 1996]).

Notice that this is not the same as claiming that we need only concern ourselves with actual intentional systems. That claim is unsatisfying if we want to know what the metaphysical conditions for mental content are. A posteriori necessary conditions hold in all possible worlds. On this view, "content' (in "intentional content") is similar to "water" and "tiger", insofar as all three refer to kinds that could have hidden or unknown essences. On this view, were Swampman to exist, he would not have intentional content, despite his having something superficially similar to it. Just so, on this view, God could not make a tiger de novo, if tigers have an essence of origin, and twin-water would not be water, if water is essentially and necessarily H20.[8]

[8]I am referring to Putnam's imaginative example, which involves a world (Twin-Earth) just like ours except that in the place of water there is a superficially indistinguishable liquid, twin-water, which is not H20. Twin-water is XYZ. Putnam's intuition, following Saul Kripke, is that

Opinion is divided as to whether this response is satisfactory. Those who find it unsatisfactory are often moved, I think, by the connections between consciousness and intentionality. While they might be willing to concede that "water" and "tiger" and the like can have hidden or unknown essences, intentionality seems different. It is intuitive, as I said before, that it would at least seem to Swampman that he was experiencing intentional mental states, and that very seeming seems to be an intentional state. There are powerful intuitions to the effect that phenomenal consciousness is intrinsic and hence shared by doppelgangers, and these intuitions about phenomenal consciousness seem to have implications for intentionality. They seem to imply that intentionality must be intrinsic and shared by doppelgangers too.

However, when assessing these intuitions it is important to keep in mind that Swampman does share with Davidson (at the moment of Swampman's inception) whatever narrow content Davidson has. This is trivial, since narrow content is defined as something that is shared by intrinsic doppelgangers, like Swampman and Davidson at the moment of Swampman's inception (along with their suitably co-ordinated Twin-Earth "twins", Matrix-trapped "twins", and so on). Those who find themselves under the sway of powerful Swampman intuitions need to consider whether this shared narrow content should placate them. It is an interesting and much discussed question what narrow content amounts to, and one we cannot enter into here. But if you think that narrow content is to some extent normative, for instance, then you should think that Swampman has some content that is to some extent normative. And this should be true of you, even if you support teleosemantics for ordinary truth-evaluative content. Teleosemantics is not inconsistent with a notion of narrow content per se. It is only inconsistent with the claim that narrow content is ordinary truth evaluative content. And this borders on the commonplace. That narrow content is not ordinary truth evaluative content was an implication of the Putnam twin- Earth thought experiments that gave rise to the literature on narrow content.

BIBLIOGRAPHY

[Biederman, 1990] I. Biederman. High level vision. In D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach (eds.), *An Invitation to Cognitive Science*, vol 2: Visual Cognition and Action. Cambridge, MIT Press, 1990.

[Boorse, 1976] C. Boorse. Wright on functions. *The Philosophical Review*, 85: 70–86, 1976.

[Brentano, 1874] F. Brentano. *Psychologie vom Empirischen Standpunkt*, vol 1, part 2, Hamburg: Meiner, 1955. (Originally published in 1874.)

[Churchland, 1995] P. Churchland. *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain. Cambridge*. MA; MIT Press, 1995.

[Churchland, 2002] P. Churchland. *Brainwise*. Cambridge, MA; MIT Press, 2002.

[Cummins, 1975] R. Cummins. Functional analysis. *Journal of Philosophy*, 72: 741–765, 1975.

[Davidson, 1987] D. Davidson. Knowing one's own mind. In *Proceedings and Addresses of the American Philosophical Association*, 60: 441–58, 1987.

our word "water" refers to water (H20) and not twin-water (XYZ), even if we do not know that water is H20. Our Twin-Earth "twins" refer to twin-water, not water, using the same vocable ("water").

[Dretske, 1986] F. Dretske. Misrepresentation. In R. Bogdan (ed.), *Belief: Form, Content and Function*, New York: Oxford, 17–36, 1986.
[Dretske, 1988] F. Dretske. *Explaining Behavior*. Cambridge, MA: Bradford, MIT, 1988.
[Dretske, 1994] F. Dretske. If you can't make one, you don't know how it works. In P. French, T. Uehling, and H. Wettstein (eds.), *Midwest Studies in Philosophy*, 19: 468–82, 1994.
[Fodor, 1990a] J. Fodor. *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press, Bradford Book, 1990a.
[Fodor, 1990b] J. Fodor. Information and representation. In P. Hanson (ed.), *Information, Language and Cognition*, University of British Columbia Press, 1990b.
[Fodor and Lepore, 1992] J. Fodor and E. Lepore. *Holism: A Shopper's Guide*. Blackwell: Oxford, 1992.
[Grice, 1957] H. P. Grice. Meaning. *The Philosophical Review*, 64: 377–388, 1957.
[Lewens, 2004] T. Lewens. *Organisms and Artifacts: Design in Nature and Elsewhere*. Cambridge, MA: MIT Press, 2004.
[Marr, 1982] D. Marr. *Vision*. Freeman Publishers, 1982.
[Matthen, 2006] M. Matthen. Teleosemantics and the Consumer. In D. Papineau and G. McDonald (eds.), *Teleosemantics*. Oxford University Press, 2006.
[Millikan, 1984] R. Millikan. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press, 1984.
[Millikan, 1989a] R. Millikan. Biosemantics. *Journal of Philosophy*, 86: 281–97, 1989a.
[Millikan, 1989b] R. Millikan. In defense of proper functions. *Philosophy of Science*, 56(2): 288–302, 1989b.
[Millikan, 1991] R. Millikan. Speaking up for Darwin. In B Loewer, G. Rey (eds.), *Meaning in Mind: Fodor and his Critics*, Cambridge, MA: Blackwell, 151–165, 1991.
[Millikan, 2000] R. Millikan. *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press, 2000.
[Neander, 1991] K. Neander. Functions as selected effects: the conceptual analysts defense. *Philosophy of Science*, 58: 169–184, 1991.
[Neander, 1995] K. Neander. Misrepresenting and malfunctioning. *Philosophical Studies*, 79: 109–141, 1995.
[Neander, 1996] K. Neander. Swampman meets Swampcow. *Mind and Language*, 11: 118–129, 1996.
[Neander, 2006] K. Neander. Content for cognitive science. In D. Papineau and G. McDonald (eds.), *Teleosemantics*, Oxford University Press, 2006.
[Papineau, 1984] D. Papineau. Representation and explanation. *Philosophy of Science*, 51: 550–72, 1984.
[Papineau, 1987] D. Papineau. *Reality and Representation*, Oxford, UK: Basil Blackwell, 1987.
[Papineau, 1993] D. Papineau. *Philosophical Naturalism*. Blackwell, Oxford, 1993.
[Papineau, 1997] D. Papineau. Teleosemantics and indeterminacy. *Australasian Journal of Philosophy*, 76: 1–14, 1997.
[Papineau, 2001] D. Papineau. The status of teleosemantics, or how to stop worrying about Swampman. *Australasian Journal of Philosophy*, 79: 279–289, 2001.
[Pietrosky, 1992] P. Pietrosky. Intentional and teleological error. *Pacific Philosophical Quarterly*, 73: 267–81, 1992.
[Rey, manuscript] G. Rey. *Intentional Inexistents in Cognitive Science*, manuscript.
[Schiffer, 1987] S. Schiffer. *Remnants of Meaning*. Cambridge, Massachusetts: MIT Press, 1987.
[Wright, 1973] L. Wright. Functions. *The Philosophical Review*, 82: 139–168, 1973.

This page intentionally left blank

INNATENESS

André Ariew

As Paul Griffiths [2002] puts it, "innateness" is associated with different clusters of related ideas where each cluster depends on different historical, cultural and intellectual contexts. In psychology innateness is typically opposed to learning while the biological opposite of innate is 'acquired'. 'Acquired' and 'learned' have different extensions. Learning is one way to acquire a character but there are others. Cuts and scratches are unlearned yet acquired; if we could acquire languages by popping a pill, then languages would be unlearned yet acquired according to the wide biological application of the term [Sober, 1998]. Further, in psychology and philosophy innateness is often associated with both "universality" (or species-specificity), and, relatedely, innate traits are often thought to be "fixed" or "unmodifiable". But, biologists recognize a range of developmental patterns that a specific trait may take. Some are universal, but others are not, as in the case of innate diseases. Some are "fixed" in the sense that once we develop them we have them for the rest of our lives; some innate diseases are like this, but others, are modifiable. Sober [1998] cites a case of an Egyptian vulture that when first confronted with an ostrich egg and a stone, will break the egg with the stone, but if the vulture repeatedly comes to find broken eggs to be empty, it will eventually stop breaking eggs. These examples lend support to Griffiths's thesis, since the concept of innateness in psychology appears to be in several ways distinct from the concept of innateness in biology.

By Griffiths's lights a reasonable conclusion to be drawn from the different meanings of "innate" is that an attempt to provide an account of innateness that crosses distinct disciplinary contexts is bound to conflate distinct biological properties and hence produce a confusing and unhelpful notion. He proposes that for each distinct context "innate" should be replaced with a term that more precisely identifies the relevant biological feature in question: "If a trait is found in all healthy individuals or is pancultural, then say so. If it has an adaptive-historical explanation, then say that" (p. 82). Griffiths's proposal has the further benefit of relieving the disciplines from invoking the "folk" concept of innateness which carries with it a false metaphysic of essentialism that Griffiths says misdescribes the identity relations in the biological world (p. 72).

Yet contrary to the spirit of Griffiths's proposal, in psychology and biology there is a strategy, let us call "biologicizing the mind", that, roughly, subsumes psychological concepts under biological models. As a broad strategy, biologicizing the mind has been quite successful. Jerry Fodor and Noam Chomsky are two well-known practitioners. Specifically on biologicizing innateness, Fodor writes: "Skin

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

color really is largely innate... much as everyone had hazily supposed. Likewise birdsong in a lot of cases; likewise the Babinsky Reflex. And it seems unlikely that the notion of innateness according to which such claims are true will prove dispensable for the larger purposes of biology" [Fodor, 2001]. Chomsky writes, "... let us consider the problem of designing a model of language acquisition... The problem is quite analogous to the problem of studying the innate principles that make it possible for a bird to acquire the knowledge that expresses itself in nest-building or in song-production" [Chomsky, 1966]. And again,

> In modern terms, that means restructuring Platonic 'remembrance' in terms of the genetic endowment, which specifies the initial state of the language faculty, much as it determines that we will grow arms not wings, undergo sexual maturation at a certain stage of growth if external conditions such as nutritional level permit this internally directed maturational process to take place, and so on [Chomsky, 1993, 519].

Both Chomsky and Fodor have clearly staked a claim in a practice that Griffiths deems unhelpful, to define the psychologist's concept of innateness in biological terms. In this essay I further explore this option. My contention with Griffiths is that it is not always true that defining a concept of innateness that crosses distinct disciplinary boundaries produces a confusing and unhelpful notion. In the case of cognitive linguistics a biologically grounded conception of innateness turns out to be extremely helpful in clarifying Chomsky's thesis that Universal Grammar is innate and particular languages are triggered rather than learned from the linguistic cues children are exposed to. Contending Griffiths's thesis in this way serves as a foil to the broader task assigned to the essays in this handbook. I will survey several biological accounts of innateness and its related concept, "triggering". I will defend a relational concept of innateness whereby innate traits are defined within a particular environmental range and refer to canalized developmental pathways. I will further argue that "triggered" traits are traits whose canalized development is initialized by a particular environmental cue.

INNATENESS AS GROWTH

Chomsky's biologicizing extends to his theory of language acquisition where, he thinks, children quite literally 'grow' languages:

> 'Language learning is not really something that the child does; it is something that happens to the child placed in an appropriate environment, much as the child's body grows and matures in a predetermined way when provided with appropriate nutrition and environmental stimulation' (p. 520).

Perhaps ascribing a character to 'growth' is all that is required to underwrite the biological concept of innate. The opposite of growth, then, is what appears by a non-growth process. On this view we could read Chomsky's work on language acquisition as suggesting that innateness refers to what grows as opposed to what is learned.

The rejection of learning models of language acquisition in favor of a growth model is the paradigm of the "biologicizing the mind" movement, but, in the context of the paradigm, an account of innateness that distinguishes growth from non-growth is unilluminating in roughly the same way that the gene/environment dichotomy has been. It is a near truism of development that every case of trait growth involves both genes and environments; genes alone or environments alone produce nothing [Lewontin, 2000]. If innateness refers to what the genes do alone then nothing is innate. Likewise, all traits grow. If we ascribe innateness to things that grow then every growing thing is innate. In all these cases we beg the further question, what sort of growth is involved? To illustrate the point, consider three different ways songbirds might develop their species-specific song (adopted from Sober [1998], who in turn cites Gould and Marler [1991]). Type 1 songbirds produce their characteristic song even if the bird is reared in silence. Type 2 birds produce their song only after sessions of call and response with a 'tutor': they attempt to mimic the song of any tutor even if the tutor happens to be a member of another species. As for Type 3 songbirds, all that is required to produce their song is contact with some song or other. They do not require a tutoring period; they require only exposure to some song. They will not respond to silence. Songs from other species or even other bird-like songs suffice to "trigger" their song capabilities. Presumably, computer generated songs suffice. Although it makes no philosophical difference to our example it might be more biologically realistic to add that Type 3 songbirds acquire their song in this unusual manner only within a 'critical' period of development. (A more realistic example will be given later — for now I seek a stark contrast afforded by the semi-fictional example). *All* three types of birdsong involve growth. If innateness means growth rather than non-growth, then innateness ascriptions will fail to pick out interesting differences between the three types of birdsong development. As a reasonable first approximation we would likely attribute "innateness" to Type 1 but not Type 2 or Type 3 songbirds since the latter two require an auditory cue for their development. But, there is a significant difference in auditory requirements between Type 2 and 3 songbirds. Invoking Chomsky's "poverty of stimulus" argument, the contact call is too impoverished to explain how Type 3 songbirds come to develop their song.

Songbird types	How species-specific song is produced
Type 1	In silence
Type 2	Requires extended call and response tutoring
Type 3	Silence is not sufficient, tutoring not necessary; all that is required is contact to some auditory cue or other

We can critique Samuels's [2002] account of innateness on similar grounds. He avers that innateness in psychology is a psychological primitive: it refers to a character whose origins are explained by biology rather than psychology. I agree that shifting the explanatory burden to biology captures the spirit of the "biologicizing the mind" movement in the cognitive sciences. Yet, as a theory of innateness, Samuels's proposal is incomplete. Perhaps (and just for the sake of argument — I think Samuels's proposal is no different than the one aforementioned proposal by Chomsky) Samuels's account could capture the difference between Type 1 and Type 2 songbirds since Type 2 (appears) to learn its song while Type 1 appears to grow it; hence we would ascribe "innateness" to Type 1 and withhold it from Type 2. But since the difference between Type 1 and Type 3 is presumably explained fully within biology rather than cutting across the psychology/biological domains, Samuels's account would ascribe "innateness" to both of them despite their developmental differences. So, Samuels's account of innateness is too weak. It fails to capture the stark contrasts in types of growth as illustrated by the birdsong example.

I propose that an appropriate account of biological innateness that applies to our psychological terms ought to be grounded in *developmental* biology, not just biology *simpliciter*. The intuition seems to comport with standard and traditional practice. What drives our interest in Plato's slave boy is that his geometric abilities seem to be something he has "in" him as opposed to being acquired by some outside influence. What distinguishes the three types of birdsong development is how they react to specific environmental cues. Yet, as I implied when I pointed out the shortcomings of the gene/environment dichotomy, if an appropriate account of innateness is to be appropriately grounded in development, there are certain facts from developmental biology that ought to be attended. Just as no trait is the product of genes alone, no system can develop without the inclusion of *some* environmental input. Type 1 songbirds, just like Plato's slave boy, require some environmental cues. Language learning, as the Chomsky quote above indicates, requires 'appropriate nutrition and environmental stimulation'. Place humans or songbirds naked on Mars and they will not develop, *period*.

So, if we want to determine whether a character is innate or acquired we ought not ask whether it develops independently of environmental interactions, period. But, we can ask a different question, *what difference does the presence or absence of certain environmental factors have on development of the trait in question*? Or, we can ask a qualitative question: what difference does fluctuating amounts of an environmental factor have on development? The second question is related to the first. If fluctuations of a certain environmental factor makes no difference to development it might be because the factor is unnecessary for development. Then again, it may be because we have not tested environmental fluctuations extreme enough to register a developmental change. Nonetheless, the lesson is the same in both cases: if we constrain our innateness ascriptions to indicate what the environment does not do to influence development then instead of uttering either biological truisms or falsehoods we might pick out interesting differences in the way

traits grow. Of course it is crucial that we pick out the right environmental factors, those that exhibit *differences* in growth patterns. We get a range of different answers when we ask what difference auditory cues make to the development of birdsong: from "nothing", as in Type 1 birdsong to an auditory cue, to "the developmental pathway is particularly sensitive to auditory inputs", as in Type 2 birdsong. In-between are developmental pathways, like Type 3 whereby the duration of the tutoring period makes no difference to the outcome — Type 3 birds' development is robust as long as it is exposed or "triggered" to some cue or other.

What is the consequence of the truism of development, that no trait develops by genes alone or environment alone, on the common folk practice of treating the innate/acquired distinction as an absolute dichotomy? One consequence supports Griffiths's eliminitivism: the truism demonstrates the inadequacy of the innate/acquired (or innate/learned) dichotomy so we should urge a change in the folk's terminology, in this case, replace the absolute dichotomy, with a more nuanced distinction regarding the degree to which a trait responds to specific cues. I favor another approach. Rather than eliminating folk usage we ought to clarify it. Perhaps innateness is dichotomously opposed to acquired in roughly the same way that a warehouse might be said to be "empty" despite it containing light bulbs and molecules. Here I am employing Dretske's conception of "relative absolutes" [Dretske, 1981]. The empty/non-empty distinction, like the innate/acquired distinction serves the pragmatic aim of picking out relative differences. Accordingly, what counts for assessing the emptiness of a warehouse might not count for assessing the emptiness of my pocket, a park, or a stadium. In each of these cases "empty" applies *relatively*. That is, the emptiness of a warehouse is relative to a certain standard, i.e., devoid of all relevant things. Likewise the innateness of Type 1 birdsong is relative to a certain set of auditory conditions as indicated in the contrast to other types of birdsong. What counts for innateness for traits in distinct environmental circumstances depends in each case on the contrast at issue. Nevertheless, for the pragmatic aim of picking out real developmental differences between organisms, innateness is dichotomous opposed to acquired in the same way that "empty" is dichotomously opposed to "non-empty": it picks out a relatively absolute term.

I have just introduced a pragmatic consideration here. Let me clarify. I am claiming that the innate/acquired distinction serves the explanatory role of picking out developmental differences between organisms (whether they be distinct types or individuals). This is consistent with my earlier claim that we cannot, on pain of a developmental truism, distinguish innate/acquired without specifying an environmental context. At this point, one might wonder, what is the value of distinguishing the three types of birds in terms of innateness/triggering/acquired? Some critics have pointed out (e.g., [Bateson and Mameli, forthcoming]) that innate ascriptions are dispensable in the light of a deep causal analysis of the developmental processes that each bird undergoes. I agree. If the question is 'how do these birds acquire their song?' then we ought to prefer a detailed causal story

rather than a rough innate/triggered/acquired distinction. But, to sometimes prefer a detailed causal story is not to undermine the value of innateness ascriptions that identify distinct developmental patterns. For instance, it is important to determine the exact nature of Type 3's trigger (a specific cue? a specific tone?) for the sake of fully explaining how Type 3 birdsong develops. But the exact nature of the trigger is not important for the sake of distinguishing Type 2 songbirds from the others (unless more information about the environmental trigger reveals that it contains a richer set of information from which growth may be sensitive).[1]

To put innate ascriptions in their place, that is, to recognize their explanatory value has a limit, it is not sufficient to dismiss the explanatory value of innate-ascriptions all together. Recently the BBC reported (May 15, 2005)[2] on a series of experiments on canaries that reveal a remarkable pattern. Typically canaries learn their song by copying adults; the tutoring process can take up to eight months. By simulating the tutoring period the experimenters managed to teach canaries to mimic non-canary computer generated songs. Yet, when injected with testosterone (simulating breeding conditions) the canaries dropped all of their learned songs and started singing *traditional* canary songs. Equally surprising, canaries that were raised without tutors sang their traditional songs when injected. The BBC report concludes:

> "counter-intuitively, although they spend a long time labouring over new songs, listening carefully, imitating and perfecting, young canaries do not actually seem to need it. Once adult, they can sing just fine without it. 'We don't have a full answer for this,' Professor Gardner told the BBC News Website".

Notice that the investigators are surprised that canary song development involves a variety of developmental patterns depending on environmental circumstances. Yet, they are even more surprised that some canaries can trigger their song from testosterone! How do they acquire *song* from such an informationally impoverished cue? Contrast triggered song development with the more ordinary developmental pattern — learning their song from a tutor. The upshot is that the discoveries of the distinctive developmental patterns that can be described in terms of "innate/acquired/triggered" is *newsworthy*. That is not to say that the investigation has ended. On the contrary, the discovery warrants further investigation of the developmental (and maybe even evolutionary) causes of the triggering phenomenon as Gardner, the lead investigor, points out.[3] Yet as a description of the phenomena to be explained innateness and triggering ascriptions are useful. They need not be misleading or wrongheaded as Griffiths insists.

[1] The debate between Chomsky and his non-innatist critics turns on the nature of the children's linguistic cues, are they mere triggers or are they rich enough to learn from?

[2] The report can be found at this website: http://news.bbc.co.uk/go/em/fr/-/2/hi/science/nature/4544777.stm. I thank Elliott Sober for bringing the article to my attention.

[3] [Gardner *et. al.*, 2005].

In this sense innate ascriptions serve a similar explanatory role as do fitness ascriptions in evolutionary biology. If we seek causes of a particular evolutionary event, say, why a population of fruitflies evolved a fuzzy thorax, then the causal details will provide a deeper explanation than an explanation that employs the fact that those flies with fuzzier thoraxes enjoyed a higher relative reproductive rate than their variants. But it does not follow that fitness explanations have no explanatory value. They serve well to describe evolutionary patterns. Further, some commentators claim that fitness explanations unify disparate evolutionary phenomena under one description [Sober, 2000; Ariew, 1996]. You do not achieve that sort of unity by citing causal details since the causal details underlying the evolution of fuzzy thoraxes are completely distinct from the causal details underlying the evolution of Saguaro cacti (for more on fitness and unifying explanations, see [Ariew, 2003]). Perhaps the same can be said about the use of the concept of innate in the cognitive sciences. In some explanatory contexts we should prefer a detailed causal story, in others, such as the case of the difference between the three sorts of developmental patterns exhibited in the bird case, we ought to prefer the blunt distinction that the "innateness"/"triggered"/"acquired" labels provide.

CHOMSKY'S POVERTY OF STIMULUS

There is a notable resemblance between Type 3 birdsong and the development of specific grammar rules in human children. All come to acquire their rich linguistic abilities despite the poor quality and quantity of the linguistic cues they receive from their linguistic communities. As I mentioned earlier, Chomsky calls the phenomenon the "poverty of stimulus" or "POS". For children, the cues are impoverished in at least two ways. First, children are exposed to a limited amount of grammar (yet their grammatical abilities are seemingly infinite). Second, the linguistic data to which a child is exposed contains errors without any indication of what distinguishes 'proper' from 'improper' grammar. The result is that any theory that postulates that a child's ability to acquire language is directly proportional to the amount of language he or she hears, e.g. a learning theory, is false.

The case of 'Simon', a child born deaf but raised by hearing parents provides a good example of POS among humans. Simon's parents had a poor grasp of the grammatical rules associated with American Sign Language (ASL) since they had to learn ASL relatively quickly once they found out about their child's condition. Despite Simon's early exposure to ASL being imperfect and crude, remarkably, Simon's own abilities to sign in ASL developed nearly 'normally' [Pinker, 1994, 39]. Even in the degraded linguistic environment whereby Simon's parents violated basic ASL grammatical rules, Simon was able to develop the 'correct' grammatical rules. The case of Simon demonstrates that despite exposure to significantly different samples of data, different children in the same linguistic community end up adopting essentially the same linguistic intuitions. Thus, it is plausible to suppose along with Chomsky that they innately possess essentially the same grammar

(more specifically, innate in the context of language cues in the environment of the learner).

Over the years, Chomsky has provided alternative theories of language acquisition that accounts for the POS phenomenon. The latest is most interesting for our purposes. Accordingly, children grow a language by utilizing an 'innate' grammar module, a 'Language Acquisition Device' (or 'LAD') that encapsulates all the possible grammar principles (a.k.a. 'universal grammar') against particular linguistic cues. Rather than learning from those cues, the cues set a 'switch' in a child's mind that leads to the adoption of the particular language of her community. Chomsky's description of the function of the 'switchbox' is worth quoting in full:

> 'The initial state of the language faculty consists of a collection of subsystems or modules... each of which is based on certain general principles. Many of these principles admit of a certain limited possibility of variation. We may think of the system as a complex network associated with a switch box that contains a finite number of switches. The network is invariant but each switch can be in one of several positions, perhaps two: on or off. But when the switches are set in one of the permissible ways, the system functions, yielding the entire infinite array of interpretations of linguistic expression' [Chomsky, 1993].

Distinguish between the development of LADs and the development of particular languages. Chomsky asserts that LADs are innate: every child has the network of switches available to them by the time they encounter linguistic cues. Loosely, the development of the LAD is like the development of Type 1 birdsong. Both develop independently of any linguistic cue. But, the development of specific grammar rules, like the development of Type 3 birdsong, is "triggered". While linguistic cues are required for their development, the cues are too impoverished to explain how the birds develop their full-blown song from them. Likewise it seems to Chomsky that the linguistic cues help set the switches from which specific grammar rules develop without the further need of linguistic output.

For an example of how linguistic cues trigger grammar growth, consider that in English, nouns, verbs, adjectives, and prepositions precede their objects. The opposite is true for Japanese. English is called a 'head-first' language, while Japanese is 'head-last'. According to Chomsky [1993, 529] simple sentences can sufficiently set the switch for children. The sentence 'JOHN ATE AN APPLE' may suffice to set the switch for English (as a head-first language), and the equivalent to 'JOHN AN APPLE ATE' sets the switch for Japanese. Chomsky concludes, "To acquire a language, the child's mind must determine how the switches are set, and simple data must suffice to determine the switch settings, as in this case" (p. 529).

Undoubtedly the analogy between linguistic and songbird development is limited. For one, Type 3 birds develop their song despite exposure to cues from birds of different species, while development of a particular language requires that the cues of the linguistic community are specific to that language. But the point of

ascribing innate/acquired to a trait is to provide a rough distinction between distinct developmental pathways. The difference concerns how the developing system reacts, in this case, to specific environmental cues, and in other cases, to amounts and quality of environmental cues. In this respect Type 3 songbirds and language development on Chomsky's theory are similar in that neither could, in the face of the POS, learn their cues from the environmental inputs. Put in another way, Chomsky's POS argument seems to be that a "triggering" model of language growth (like Type 3 birdsong) better predicts child development given the POS than a learning model (like Type 1 birdsong).

To sum up so far: I began the discussion of birdsong, language development, and POS with a statement of what, from a developmental biological point of view, could possibly ground the distinction between "innate" and "acquired". I suggested that the distinction depends on what certain environmental cues can or cannot do to effect growth of the trait in question. Auditory cues have no effect on the development of Type 1 birdsong, yet, in contrast, species-specific auditory cues are required for development of Type 2 birdsong. Type 3 songs require some auditory cue but the effect of the cue does not serve (as it does for Type 2 birdsong) to shape the end state. Rather, in the case of Type 3 songbirds the auditory cue serves as a "trigger". If Chomsky is right, we have a similar situation for language development. The development of LAD does not depend on linguistic cues while the development of specific grammar rules require some set of rather specific cues. Yet, in the face of POS, the cues appear to serve as a "trigger" to set switches of an LAD switchbox rather than as a source from which the languages are shaped or "learned".

CANALIZATION AND THE EPIGENETIC LANDSCAPE

Next, we need a general account in biology that adequately captures the following intuitions: a) that innateness means more than unlearned, b) that a biological conception refers to biological development, and c) an adequate account captures relevant developmental differences between traits that get their trait independent of linguistic cues (like type 1, and LAD) and those that require some linguistic cue or other, whereby the cue is too impoverished to explain the output (like Simon, type 3 and the head first/head last grammar rules).

Elliott's Sober's proposal [1998, 795] is a good start. "A phenotypic trait is innate for a given genotype if and only if that phenotype will emerge in all of a range of developmental environments". In short, innateness amounts to phenotypic invariance across a range of environmental conditions. If what I've argued earlier is correct, that innate ascriptions on the biological model should indicate what the environment can or cannot do to affect the development of a trait, then Sober's invariant account is on the right track. In the context of auditory cues, Type 1 birdsong capabilities are more invariant than Type 2 or Type 3 birdsong, because Type 1 birdsong emerges in an extra environment, where auditory cues are absent. The type that develops the trait in the absence of the condition in the environment

would be said to be more invariant, and, on Sober's account, more innate.

Three related problems emerge from Sober's account. First, Sober's account leaves open the question how does one distinguish the ranges of environmental variation that are relevant to assess the innateness of a trait and the ranges that are not? Some take this to be a serious shortcoming [Mameli and Bateson, 2005; Cowie, 1998; Prinz, 2002]. Every developing trait will be sensitive to some environmental variation and insensitive to others, hence, a consequence of Sober's account is that a trait is innate with respect to some environmental conditions and non-innate with respect to others. Sober admits this *lacunae*, though he questions whether there is a uniquely correct answer to what counts as the appropriate environmental range. He avers, "maybe the range is determined pragmatically. It is difficult to see how the latter conclusion can be evaded" (p. 795). Yet, there is a sense in which despite the pragmatic or explanatory aims, certain environmental conditions distinguish real developmental differences. Let me say more here. I have argued that innate ascriptions in biology serve to distinguish between how various developmental systems react to specific environmental cues. The difference between the three types of birdsong is determined by how they react to auditory cues. Further, as the canary example illustrates, individuals possess a variety of potential developmental outcomes. When canaries contain some level below a threshold of testosterone in their systems their song development requires a tutoring period. But, when the need to procreate becomes urgent (when testosterone levels are above the threshold) the same canaries develop their song without the need for a tutoring period. The lesson learned here is that there are real developmental differences between organisms and even contingently within organisms with respect to how they will react to certain environmental cues. Ever since the work of C. H. Waddington in the 1950s, developmental biologists have recognized that context dependency is an important feature of developmental systems. A developmental system responds to certain environmental cues by changing its expression patterns. Waddington called this the phenomenon of the "reactive genome". In the contemporary literature the phenomenon is termed "tertiary induction" [Gilbert, 2004, 350]. If picking out these real differences is the point of innate ascriptions, then perhaps there is a principled answer to the question what counts as an appropriate environmental range, namely, the ranges in which the expression patterns become insensitive to environmental perturbations, like in the case of testosterone induced canaries, or Type 1 songbirds. Waddington called the process of buffering against environmental cues "canalization". Perhaps we should amend Sober's account accordingly and identify innateness with canalization [Ariew, 1996; 1999].

The canalization amendment is significant for another reason, it solves a second problem with Sober's account. Distinguish between two reasons why the trait appears invariantly in an environmental range: the first, because an environmental condition is developmentally required yet is found everywhere the system develops; the second, because the system develops *independently* of the environmentally condition. Innateness should be identified with the second sort of invariance, not the first. Yet, Sober's account fails to recognize the difference (the following is from

[Ariew, 1999]). For example, consider the intestinal bacterium, *clostrium difficile* (*c. diff.*), that we humans invariably acquire in the food and drink we consume (the example comes from [Wendler, 1996]). *C. diff* is invariantly acquired, that is, it emerges in all of a range of human environments. Hence, on Sober's account the presence of *c. diff.* is innate. But its invariance is due to a specific environmental condition that is everywhere humans are since it is present in the food and water that humans drink. Hence, intuitively, *c. diff.* is acquired, not innate. Note, the appropriate environmental range that picks out the difference is only conceptually possible, the environment where humans develop without food or water. If human stomachs contained *c. diff.* in environments where no *c. diff.* is present then likely humans have a canalized developmental pathways to insure its presence in the stomach. If the idea that *c. diff.* development is canalized sounds too improbable, think about Type 1 songbirds developing — the lesson is the same — the fact that their song develops independent of auditory cues suggest that Type 1 birds song development is canalized or buffered against the absence or presence of auditory cues.

Samuels [2002] offers a cognitive example that serves to illustrate the same ambiguity of invariance accounts. Presumably the belief 'water is wet' is learned from our interaction with water. But, since water is everywhere humans are the belief emerges invariably in human environments. Samuels believes his counter-example warrants a rejection of all developmental invariance accounts. Yet, on the face of it, canalization picks out the appropriate distinction and properly identifies innateness with development that is independent of the environmental cue in question (see [Collins, 2004] for further discussion).

Canalization appeases a third worry I have of Sober's invariance account. Consider the three birds again. On the invariance account the difference between the birds is depicted merely as a matter of degree depending on the number of environments where song emerges. On the one end of the continuum (towards "more innate") is Type 1 birdsong, Type 2 is at the other end, and Type 3 is somewhere in the middle. Given that the three birds are being compared in a common environment, Type 3 songbirds are not just a matter of degree distinct, they are distinct in *kind*.[4] While unlike Type 1, Type 3 songbirds *require* an auditory cue. Yet, the nature of the relationship between the developmental system and the auditory cue is wholly unlike that of Type 2 songbirds. Type 3 as opposed to Type 2 birdsong exemplifies the POS since Type 3 are able to exhibit their species-specific song even when the auditory cue is so degraded that it could not possibly learn from it. On my view, the concept of canalization, and its related concept, the "epigenetic landscape" accounts for the relevant differences: Type 1 song development is canalized, Type 2 is not, Type 3's canalized development is "triggered" or on some auditory cue or other. To fill out the idea, let me give some background information on canalization and the epigenetic landscape from which the concept of canalization is drawn.

[4]In what follows I depart a bit from Ariew [1996; 1999] where canalization is merely a matter of degree.

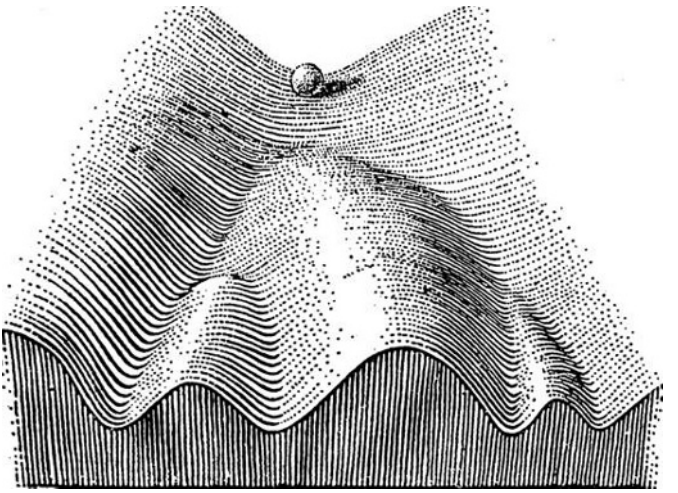


Figure 1. (taken from [Waddington, 1957])

C. H. Waddington introduced canalization to the developmental biology literature in the 1950s as a feature of his simple model of development, called the 'epigenetic landscape' (see especially Waddington [1957]). Waddington was motivated to explain a peculiar feature of development experienced by vast number of species: individuals tend to develop into one of a few distinct body types despite the great environmental and genetic variation between them.

In Figure 1, the ball represents a developmental system at some stage. The branching system of 'canals' (or Waddington's term, 'creodes') represent possible pathways the system might encounter within an environmental range. Depending on some number of environmental or genetic factors, the developmental system will be 'nudged' down one or another 'canal'. The depth of the canals represent the degree to which the development of the end state is hard to change, disrupt or impair in the face of further genetic or environmental perturbations. Once the system reacts to the initiating cue the development of the end state is more or less 'fated', barring extreme environmental conditions.

There are two sorts of canalization: the degree to which a system is so buffered against *genetic* perturbations is the degree to which it is 'genetically canalized'. The degree to which a system is so buffered against *environmental* perturbation is the degree to which it is 'environmentally canalized' (for more details see [Waddington, 1957]). For our purposes, since innateness is meant to reflect what the environment does not do to effect development of the end state, I propose innateness is associated with environmental and not genetic canalization. From here on out when I say "canalization" I am referring to "environmental canalization".

Looking at the epigenetic landscape in Figure 1 there appears two ways to characterize the extent to which a system is canalized. A developmental system

may be canalized to a greater degree if it produces its end state across a greater environmental range than the other. That is, one trait is more canalized than another if the one is more developmentally invariant than the other. It is this sense of canalization that accounts for the invariance effect that is central to Sober's account of biological innateness. But better than a mere invariance account the effect is grounded in a real developmental pathway.

Canalized systems might also be compared by the degree to which fluctuating environmental conditions affect development of a trait already in a canal or "chreode" (this is represented in Figure 1 by the highness of the canal walls). It is this consideration, how fluctuating environmental conditions affect development that is the extra feature missing in Sober's invariance account. Consider Figure 2:

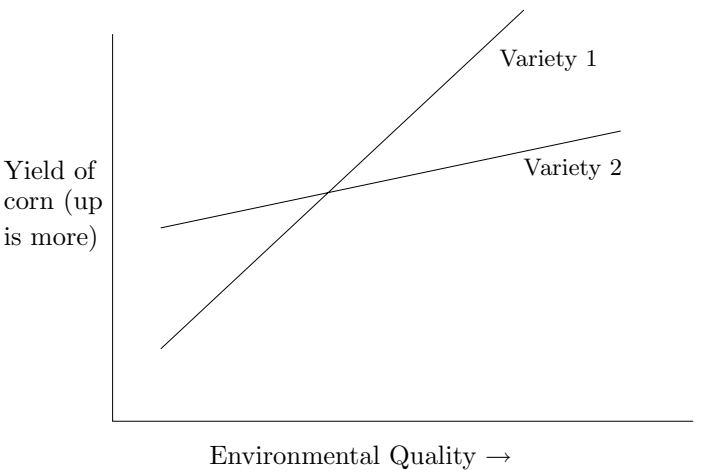


Figure 2.

Variety 2 corn plants produce their yields more robustly than variety 1 corn plants. Variety 2 yields are less sensitive than Variety 1 yields to fluctuations in environmental quality. In this case 'fluctuations' refer to varying conditions *already present* in the environment. Indeed, the factors that determine environmental quality need to be present for any corn plant to grow; no plant will grow without some amount of soil.

The difference between the two varieties of corn plant yields is similar in kind to the difference between Type 2 and 3 birdsong where all *require* an auditory cue while each type reacts differently to the environmental quality (auditory cue). Type 3 songbirds tolerate a larger range of auditory cues, including completely degraded ones. That is why we might be tempted to distinguish the relation between development and the cue in this case as "triggered" growth. In contrast, Type 2 songbirds are sensitive to the quality of the auditory cues. Type 1 songbirds would exhibit a completely flat line since it is insensitive to any amount of fluctuation. Yet what is distinctive about Type 1 songbirds is not represented in the

graph, more than being insensitive to fluctuations, its development is completely independent of auditory cues.

Mameli and Bateson usefully distinguish between "developmental canalization" and "post-developmental canalization". Accordingly, "a phenotype P is developmentally canalized if an evolved mechanism M exists to ensure that P develops in the face of certain perturbations, and post-developmentally canalized if an evolved mechanism M exists to ensure that P is not modified by the occurrence of certain events after its development is complete" (p. 18). The difference between developmental canalization and post-developmental canalization is another feature of the epigenetic landscape. Some developmental systems are sensitive to environmental cues only during a critical stage of development. According to Chomsky, there is a critical stage for the development of languages. I built a critical period into the description of Type 3 songbirds. The concept that the developmental landscape features forks in the road depending upon certain environmental cues given at crucial stages is an important insight for "eco-devo" a burgeoning field of development. Examples abound: the sex of snapping turtle depends on temperature at the embryo stage, at one temperature the turtle embryo becomes male, at another it becomes female. Ant larvae develop into either sterile workers or fertile queens depending on the diet they are fed (see [Gilbert, 2004]). We wouldn't want to say that queen development is canalized against diet, since whether a larva will develop into a queen or sterile worker is particularly sensitive to diet. But, once the special diet is fed and all other environmental conditions are held constant, then queen development rolls down a canalized pathway with very high walls on the epigenetic landscape. The distinction comports well with innate ascriptions. The difference between queens and sterile workers is not wholly innate, it would seem, because the diet plays the crucial difference. But 'acquired' does not seem to capture the appropriate distinction either. The cue seems too impoverished to explain the radical difference between queens and sterile workers given a particular diet at a particular time. And furthermore, once the special diet is introduced, queens invariantly develop. 'Acquired' implies a relationship between larva and diet that would resemble something closer to Variety 1 corn plants — the longer the diet the more likely the larva develops into a queen. Instead the developmental profile of queens is more like Variety 2 corn plants. Once diet is introduced, queen development is robust. We could say that the difference is a matter of degree (as Sober might say), but "canalization that is triggered on diet" seems to capture the innateness intuitions better.

I favorably mentioned Mameli and Bateson's distinction between "developmental canalization" and "post-developmental canalization" but the difference between Mameli and Bateson's approach and mine concerns their insistence that innateness refers to evolved canalized mechanisms. Presumably they mean "evolved by natural selection" to indicate a pathway that is advantageous to the individual. I agree with Mameli and Bateson that a virtue of the canalization account of innateness is that it can be assimilated to the idea of natural selection (see also [Ariew, 1996; 1999]). Yet, making adaptation a necessary condition for canalization grounded

innateness is too restrictive. As Mameli and Bateson themselves point out some fitness-detrimental diseases develop along canalized pathways. Yet, since disease development does not evolve for the benefit of the individual victims, Mameli and Bateson think that disease development is a counter-example to the canalization account. But, the counter-example is effective in the context of their *desiderata* that includes as a necessary condition for a trait to be innate that it is the product of natural selection.

I reject the evolutionary desiderata on independent grounds. Consider a distinction made famous by Ernst Mayr between proximate causes and ultimate causes whereby the proximate causes explain (among other things) the development of the trait in question. Ultimate causes are cited to explain how developmental systems of the type in question have come to evolve. Presumably, the development patterns of all three songbirds evolved, and perhaps they are all adaptations (evolved by natural selection). Regardless, the most important feature of the distinct birdsong types is best invoked by their proximate causes — each bird presents a distinct sort of developmental pattern. Innate ascriptions pick out those differences. Whether our language abilities are the products of natural selection or not is one question, but it isn't the relevant question linguists are asking when they want to know what sort of developmental processes are involved when children acquire languages.

Mameli and Bateson were motivated to include "evolution" as a condition for innateness on the canalization model to solve a problem of providing a principled distinction between what counts as the relevant environmental range and what does not (p. 18). Mameli and Bateson think the lack of a principled distinction is a problem for invariance accounts like Sober's. (They wrongly interpret my 1999 account as an invariance account despite my explanation to the contrary.) What I have been arguing with the disease example is that there ought not be any further condition for a canalized based account of innateness that canalized buffering mechanisms have to benefit the individual.

In summary, let us ask what do we learn about innateness from a developmental point of view? In general Waddington's concepts provide us with a cluster of related ideas useful for making the appropriate distinctions along the innate/acquired spectrum that can be used to understand innate ascriptions in the cognitive sciences. The discovery that developmental systems have an ability to buffer development against environmental perturbations to ensure the production of an end state suggests that some innate ascriptions in the cognitive science might be biologically grounded. I think much of contemporary cognitive science, especially those moved by the "biologicizing the mind" movement, implicitly or explicitly employ the concept of innateness as it relates to canalization or some aspect of the epigenetic landscape, though I have focussed my demonstration on Chomskian linguistics.

On first approximation, the epigenetic landscape describes the relation between the three types of birdsong development. Type 1 birdsong development is highly environmentally canalized across auditory cues, meaning that no linguistic cue or perturbations in the auditory signals would prevent the song from developing. In

fact, the development of Type 1 songs develop independently from any auditory cue. To translate this into an account of innateness, we might say that Type 1 birdsong is innate with respect to auditory cues. Type 2 birdsong is plastic, meaning that its development is not canalized with respect to auditory cues. In comparison, Type 3 birdsong development is contingent upon the presence of auditory cues as some stage of development. The epigenetic fork in the road represents the status of a Type 3 songbird. If an auditory cue is encountered, then its subsequent song development is canalized, otherwise, not. The canary song discussed in the BBC report has an unusual epigenetic landscape because it seems that across a wide range of environmental conditions song development is relatively plastic rather than canalized. But, the introduction of testosterone at any point in the developmental stage is enough to initiate a canalized pathway. Waddington achieved the same result with fruit flies and an unusual environmental condition. Most fruit flies develop one set of wings and a single thorax. But, some when exposed to ether at a crucial stage of development, some flies responded with a second thorax and a second pair of wings. The lesson here is that while across "normal" environmental conditions, some features might be highly canalized, but development might at the same time be sensitive to unusual or specific environmental cues such that their presence is enough to trigger another canalized pathway.

In Cowie's critique of the canalization account of innateness she writes, "the arguments from the poverty of the stimulus nativists use to defend their position do not in fact entail anything about the degree of plasticity possessed by the processes responsible for our acquisition of ideas and beliefs. For the fact that the outputs of learning might be thoroughly underdetermined by the environmental information (as poverty of the stimulus arguments contend) is quite consistent with any amount of plasticity in the learning process itself" (p. 46). Yet, if innateness is canalization as opposed to mere developmental invariance then the developmental response to an environmental trigger does in fact tell us a lot about the degree to which innateness is either innate, triggered, or acquired. Notice, Chomsky's "switchbox" model of grammar adoption dovetails nicely with the epigenetic landscape. The adoption of 'head-first' languages are "triggered" or phenotypically switched by a few linguistic cues. Once the triggering environmental cue is encountered, development of one or the other pathway is relatively unaffected by the presence or absence (or poor quality) of further linguistic cues. Perhaps post-trigger development proceeds independently of linguistic cues. If so, we would say that post-trigger development of 'head-first' or 'head-last' grammar is innate (*simpliciter*) across linguistic cues. Otherwise we would say that it is simply to some degree canalized. Either way, compared to learning models of grammar acquisition Chomsky's switchbox model predicts that the development of specific grammar rules is relatively robust. As evidence by the POS the development of specific grammar rules appears relatively unaffected by fluctuations of quality and quantity of linguistic cues, suggesting that the development of grammar rules is to some degree canalized, though it is not innate since grammar rules require certain linguistic cues.

As Chomsky has pointed out to me, in both cognitive science and biology it is useful as a first approximation to distinguish between a case where what is innate specifies in a significant way the form of the outcome from a case in which what is innate is a particular set of procedures to apply to external inputs without further indication of the outcome. Chomsky takes the former, what he has called "Rationalism" (see [1967]) to be a hallmark of canalization. In this essay I have attempted to associate the concept of innateness in the cognitive sciences, specifically in the literature surrounding Chomsky's theory of language acquisition, with the biological concept of canalization. In the spirit of Griffiths's proposal that innateness is defined loosely around a cluster of biological principles, I have shown how the canalization concept serves as a first approximation to determine differences between developmental pathways, especially on how each might react to a set of environmental cues. I proposed that the innate/acquired dichotomy can be preserved in developmental cases where canalization ensures the development of an end state even when a particular environmental cue is not present. I proposed that "triggered" traits are environmental phenotypic switches that initiate canalized pathways.

Fodor writes: 'A lot of people have Very Strong Feelings about what concepts are allowed to be innate... Almost everybody is prepared to allow RED in, and many of the liberal-minded will also let in CAUSE or AGENT... But there is, at present, a strong consensus against, as it might be, DOORKNOB or CARBURETTOR. I have no desire to join in this game of pick and choose since, as far as I can tell, it hasn't any rules' [Fodor, 1998, 28]. I hope that I have shown that indeed there are rules.

ACKNOWLEDGEMENTS

For comments on the latest draft, I would like to thank David Adger, Patrick Bateson, Ron Mallon, Matteo Mameli, and Chris Stephens.

BIBLIOGRAPHY

[Ariew, 1996] A. Ariew. Innateness and canalization. *Philosophy of Science*, 63 (Proceedings): S19–S27, 1996.

[Ariew, 1999] A. Ariew. Innateness is canalization: A defense of a developmental account of innateness. In V. Hardcastle (ed.), *Biolology Meets Psychology: Conjectures, Connections, Constraints*. Cambridge, MA: MIT Press, 1999.

[Ariew, 2003] A. Ariew. Ernst Mayr's 'ultimate/proximate' distinction reconsidered and reconstructed. *Biology and Philosophy*, 18: 553–565, 2003.

[Chomsky, 1966] N. Chomsky. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper and Row, 1966.

[Chomsky, 1980] N. Chomsky. Rules and representations. *The Behavioral and Brain Sciences*, 3: 61, 1980.

[Chomsky, 1993] N. Chomsky. On the nature, use, and acquisition of language. In A. Goldman (ed.), *Readings in Philosophy and Cognitive Science*. Cambridge, MA: MIT Press, 1993.

[Chomsky, 2003] N. Chomsky.: Reply to Gopnik. In L. Antony and N. Hornstein (eds.), *Chomsky and his critics*. Oxford: Blackwell, 2003.

[Collins, 2005] J. Collins. Nativism: In defense of a biological understanding. *Philosophical Psychology*, 18(2): 157–177, 2005.
[Cowie, 1988] F. Cowie. *What's Within*. Cambridge, MA: MIT Press, 1998.
[Dretske, 1981] F. Dretske. The pragmatic dimension of knowledge. *Philosophical Studies*, 40: 363–78, 1981.
[Fodor, 1975] J. Fodor. *The Language of Thought*. Cambridge, MA: Harvard University Press, 1975.
[Fodor, 1981] J. Fodor. *Representations*. Cambridge, MA: MIT Press, 1981.
[Fodor, 1998] J. Fodor. *Concepts*. New York, NY: Oxford U. Press, 1998.
[Fodor, 2001] J. Fodor. Doing without what's within: Fiona Cowie's critique of nativism. *Mind*, 110(437): 99–148, 2001.
[Gardner *et al.*, 2005] T. J. Gardner, F. Naef and F. Nottebohm. The acquisition an dreporgramming of a bird's learned song. *Science*, 308(5724): 1046–1049, 2005.
[Gilbert, 2004] S. Gilbert. Ecological developmental biology: Developmental biology meets the real world. *Russian Journal of Developmental Biology*, 35(6): 346–357, 2004.
[Godfrey-Smith, 1995] P. Godfrey-Smith. *Complexity and the Function of Mind in Nature*. New York, NY: Cambridge U. Press, 1995.
[Gould and Marler,, 1991] J. Gould and P. Marler. Learning by instinct. In D. Mock (ed.) *Behavior and Evolution of Birds*. San Francisco, CA: Freeman, pages 4–19, 1991.
[Griffiths, 2002] P. Griffiths. What is innateness? *The Monist*, 85(1): 70–85, 2002.
[Lewontin, 2000] R. Lewontin. *The Triple Helix: Gene, Organism and Environment*. Cambridge, MA: Harvard U. Press, 2000.
[Mameli and Bateson, 2005] M. Mameli and P. Bateson. Innateness and the sciences. *Biology and Philosophy*, 2005 (forthcoming).
[Oyama, 1985] S. Oyama. *The Ontogeny of Information: Developmental Systems and Evolution*. Cambridge, MA: Cambridge University Press, 1985.
[Pinker, 1994] S. Pinker. *The Language Instinct*. New York, NY: W. Morrow and Co., 1994.
[Prinz, 2002] J. Prinz. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press, 2002.
[Samuels, 2002] R. Samuels. Nativism in cognitive science. *Mind and Language*, 17(3): 233–265, 2002.
[Sober, 1998] E. Sober. Innate knowledge. In E. Craig, L. Floridi (eds.) *Routledge Encyclopedia of Philosophy*. Cambridge: Routledge, pages 794–797, 1998.
[Sober, 2000] E. Sober. *Philosophy of Biology*. Westview Press, 2000.
[Sterelny, 1989] K. Sterelny. Fodor's nativism. *Philosophical Studies*, 55: 119–141, 1989.
[Waddington, 1957] C. H. Waddington. *The Strategy of the Genes*. London, Allen & Unwin, LTD, 1957.
[Wendler, 1996] D. Wendler. Locke's acceptance of innate concepts. *Australasian Journal of Philosophy*, 74(3): 467–483, 1996.

ARTIFICIAL LIFE

Mark A. Bedau

Contemporary artificial life (also known as "ALife") is an interdisciplinary study of life and life-like processes. Its two most important qualities are that it focuses on the essential rather than the contingent features of living systems and that it attempts to understand living systems by artificially synthesizing extremely simple forms of them. These two qualities are connected. By synthesizing simple systems that are very life-like and yet very unfamiliar, artificial life constructively explores the boundaries of what is possible for life. At the moment, artificial life uses three different kinds of synthetic methods. "Soft" artificial life creates computer simulations or other purely digital constructions that exhibit life-like behavior. "Hard" artificial life produces hardware implementations of life-like systems. "Wet" artificial life involves the creation of life-like systems in a laboratory using biochemical materials.

Contemporary artificial life is vigorous and diverse. So this chapter's first goal is to convey what artificial life is like. It first briefly reviews the history of artificial life and illustrates the current research thrusts in contemporary "soft", "hard", and "wet" artificial life with respect to individual cells, whole organisms, and evolving populations. Artificial life also raises and informs a number of philosophical issues concerning such things as emergence, evolution, life, mind, and the ethics of creating new forms of life from scratch. This chapter's second goal is to illustrate these philosophical issues, discuss some of their complexities, and suggest the most promising avenues for making further progress.

1 HISTORY AND METHODOLOGY

Contemporary artificial life became known as such when Christopher Langton coined the phrase "artificial life" in the 1980s. Langton described artificial life as a study of life as it could be in any possible setting and he organized the first conference that explicitly recognized this study [Langton, 1989].

The intellectual roots of contemporary artificial life grow back to the first half of the twentieth century, and the two deepest roots reach to John von Neumann and Norbert Wiener. Von Neumann [1966] designed the first artificial life model (without referring to it as such) when he created his famous self-reproducing, computation-universal cellular automata.[1] Von Neumann tried to understand the

[1] A cellular automaton is a regular spatial lattice of "cells," each of which can be in one of a finite number of states. The lattice typically has 1, 2, or 3 spatial dimensions. The state of

Handbook of the Philosophy of Science. Philosophy of Biology
Volume editors: Mohan Matthen and Christopher Stephens
General editors: Dov M. Gabbay, Paul Thagard and John Woods
© 2007 Elsevier B.V. All rights reserved.

fundamental properties of living systems, especially self-reproduction and the evolution of complex adaptive structures, by constructing simple formal systems that exhibited those properties. At about the same time, Wiener [1948] started applying information theory and the analysis of self-regulatory processes (homeostasis) to the study of living systems. The abstract constructive methodology of cellular automata still typifies much artificial life, as does the abstract and material-independent methodology of information theory.

Artificial life has also been influenced by developments in traditional disciplines. Wet ALife clearly grows out of work in molecular biochemistry on the origin of life, and artificial life in general clearly benefits from a wealth of information about life on Earth. In addition, some models originally devised for specific biological phenomenon have subsequently been adopted and explored for other purposes by artificial life, e.g., the random Boolean networks originally introduced by Kauffman as a model of gene regulation networks.[2] Physics and mathematics, especially statistical mechanics and dynamical systems, have contributed the method of constructing simple model systems that have broad generality and permit quantitative analysis. Furthermore, the use of cellular automata as exemplars of complex systems [Wolfram, 1994] directly led to contemporary artificial life.

Much of the early work on artificial life was showcased at the Santa Fe Institute, an interdisciplinary research institution that helped put the study of complex systems on the map. Complex systems are composed of many elements that are simultaneously interacting with each other. Those in which the rules governing the elements are reshaped over time by some process of adaptation or learning are complex adaptive systems [Holland 1975/1992; 1995]. Artificial life focuses specifically on those complex systems that involve life, and these typically involve adaptation and learning.

Though artificial life differs from artificial intelligence, the two are connected through ALife's deep roots in computer science, especially artificial intelligence (AI) and machine learning. Notable here are John Holland's pioneering investigations of genetic algorithms [1975/1992].[3] The subjects of AI and artificial

each cell in the lattice is updated simultaneously in discrete time steps. Each cell is a finite state machine that outputs the next state of the cell given as input the states of the cells within some finite, local neighborhood of the lattice. Typically all cells in the lattice are governed by the same finite state machine, which typically is deterministic.

[2]Random Boolean networks consist of a finite collection of binary (ON, OFF) variables with randomly chosen input and output connections. The state of each variable at each step in discrete time is governed by some logical or Boolean function (AND, OR, etc.) of the states of variables that provide input to it. The network is started by randomly assigning states to each variable, and then the connections and functions in the network determine the successive state of each variable. Since the network is finite, it eventually reaches a state it has previously encountered, and from then on the network will forever repeat the same cycle of states. Different network states can end up in the same state cycle, so a state cycle is called an attractor.

[3]The genetic algorithm is machine learning technique loosely modeled on biological evolution. It treats learning the solution to a problem as a matter of competition among candidate problem solutions, with the best candidate solutions eventually winning. Potential solutions are encoded in an artificial chromosome, and an initial population of candidate solutions is created randomly. The quality or "fitness" of each solution is calculated by application of a "fitness function."

life overlap, since living and flourishing in a changing and uncertain environment requires at least rudimentary intelligence. Their methodologies are also similar, since both study natural phenomena by simulating and synthesizing them.

Nevertheless, there is an important difference between traditional symbolic AI and artificial life. Most traditional AI models are top-down-specified serial systems involving a complicated, centralized controller that makes decisions based on access to all aspects of global state. The controller's decisions have the potential to affect directly any aspect of the whole system. On the other hand, many natural living systems exhibiting complex autonomous behavior are parallel, distributed networks of relatively simple low-level "agents" that simultaneously interact with each other. Each agent's decisions are based on information about only its own local environment, and its decisions directly affect only its own local environment.

ALife's models characteristically follow this example from nature. The models themselves are bottom-up-specified parallel systems of simple agents interacting locally. The models are repeatedly iterated and the resulting global behavior is observed. Such lower-level models are sometimes said to be "agent-based" or "individual-based." The whole system's behavior is represented only indirectly. It arises out of interactions among directly represented parts ("agents" or "individuals") and their physical and social environment. This decentralized architecture shares important similarities with some newer trends in AI, including connectionism [Rumelhard and McClelland, 1986], multiagent AI [Rich and Knight, 1991], and evolutionary computation [Holland, 1975/1992; Mitchell, 1996].

An accurate and detailed sense of artificial life's central aims can be found in the unabashedly long-term grand challenges framed by the organizers of Artificial Life VII, the International Conference on Artificial Life that occurred at the new milennium [Bedau *et al.*, 2000]. The challenges fell into three broad categories concerning life's origin, its evolutionary potential, and its connection to mind and culture.

How does life arise from the non-living?

1. Generate a molecular proto-organism *in vitro.*
2. Achieve the transition to life in an artificial chemistry *in silico.*
3. Determine whether fundamentally novel living organizations can arise from inanimate matter.
4. Simulate a unicellular organism over its entire lifecycle.

For example, if the problem is to find the shortest route between two cities and a candidate solution is a specific itinerary, then the fitness function might be the sum of the distance of each segment in the itinerary and a solution's fitness is proportional to the reciprocal of its total distance. In effect, the fitness function is the "environment" to which the population adapts. A candidate solution's "genotype" is its chromosome, and its "phenotype" is its fitness. On analogy with natural selection, lower fitness candidates are then replaced in the population with new solutions modeled on higher fitness candidates. New candidates are generated by modifying earlier candidates with "mutations" that randomly change chromosomal elements and "crossover" events that combine pieces of two chromosomes. After reproducing variants of the most fit candidates for many generations, the population contains better and better solutions.

5. Explain how rules and symbols are generated from physical dynamics in living systems.

What are the potentials and limits of living systems?

6. Determine what is inevitable in the open-ended evolution of life.
7. Determine minimal conditions for evolutionary transitions from specific to generic response systems.
8. Create a formal framework for synthesizing dynamical hierarchies at all scales.
9. Determine the predictability of evolutionary manipulations of organisms and ecosystems.
10. Develop a theory of information processing, information flow, and information generation for evolving systems.

How is life related to mind, machines, and culture?

11. Demonstrate the emergence of intelligence and mind in an artificial living system.
12. Evaluate the influence of machines on the next major evolutionary transition of life.
13. Provide a quantitative model of the interplay between cultural and biological evolution.
14. Establish ethical principles for artificial life.

Some areas of artificial life are missing from the list, notably hard artificial life. This is simply an historical accident of who attended Artificial Life VII.

2 THREE ILLUSTRATIONS OF CONTEMPORARY ARTIFICIAL LIFE

Life exhibits complex adaptive behavior at many different levels of analysis, ranging from individual cells to whole organisms, and even to evolving ecologies. One can get a concrete feel for contemporary artificial life by considering a few illustrations of soft, hard, and wet artificial life. These examples illustrate artificial life's broad interdisciplinary nature, its synthetic methodology, and its concern with understanding the essential properties of living systems.

Artificial cells

The holy grail of "wet" artificial life is to create artificial cells out of biochemicals [Bedau *et al.*, 2000; Rasmmssen *et al.*, 2007]. Such artificial cells would be microscopic autonomously self-organizing and self-replicating entities built from simple organic and inorganic substances [Rasmussen *et al.*, 2004]. Although artificial, for

all intents and purposes they would be alive, for they would maintain themselves, spontaneously organize and repair themselves, and adapt in an open-ended fashion to environmental contingencies.

There are two main motivations behind this research. One is pure science. If one could make artificial cells from scratch, especially using materials or methods that are not employed by nature, one would have dramatic proof of the possible molecular foundations of living systems. Artificial cells also have a broad practical appeal. Natural cells are much more complicated than anything yet produced by man, and many people believe that the next watershed in intelligent machines depends on bridging the gap between non-living and living matter [Brooks, 2001]. So, making artificial cells that organize and sustain themselves and continually adapt to their environment would open the door to future technologies with the impressive capacities of living systems.

What will artificial cells do? The initial functionality of these machines will be simply to move through a fluid and process chemicals. To do this flexibly and resiliently involves solving the basic functions of self-maintenance, autonomous control of chemical processing, autonomous control of mobility, and self-replication. Artificial cells will simultaneously solve these tasks by integrating an artificial metabolism with the means of growth and self-reproduction, and localizing these chemical systems by producing some container. Thus, artificial cells will have biochemical systems that construct and repair the system's container (e.g., cell walls), systems that copy the information-bearing molecules that encode and direct cellular processes (genes), and systems that synthesize the materials for cellular self-assembly and regeneration (a metabolism). All life in nature depends on the coordination of these three processes. The first artificial cells will have extremely simple versions of them.

Nobody has yet created an artificial cell, but research toward this goal is actively under way. Two main approaches are being pursued. Human genome pioneer J. Craig Venter and Nobel Prize winner Hamilton Smith recently publicized their intention to create a partly man-made artificial cell, with support from the US Energy Department [Gillis, 2002]. Venter and Smith are using the top-down strategy of simplifying the genome of the simplest existing cell with the smallest genome: *Mycoplasma genitalium* [Fraser *et al.*, 1995; Hutchison *et al.*,1999]. This top-down approach has the virtue that it can simply borrow the biological wisdom embodied in *Mycoplasma* biochemistry. It has the corresponding disadvantage that its insights will be limited by various contingencies of *Mycoplasma's* evolution.

The other approach to making artificial cells is bottom up: to build more and more complex physiochemical systems incorporating more and more life-like properties. Szostak, Bartell, and Luisi [2001] and Pohoril and Deamer [2002] describe bottom-up strategies that are strongly inspired by the lipid bilayer membranes and nucleic acid chemistry found in existing cells. Lipid vesicles have been shown to grow and reproduce in the laboratory [Walde *et al.*, 1994; Menger and Angelova, 1998]. The main challenge of this bottom-up strategy is that there is no known chemical path for synthesizing DNA or RNA that is sufficiently complex to en-

code the minimal molecular functions needed by such artificial cells. Rasmussen *et al.*,[2003] have proposed a simpler and much less natural bottom-up approach in which PNA chemistry [Nielsen *et al.*,1991] replaces RNA chemistry and lipid micelles replace vesicles.

Autonomous agents

Much work in artificial life at the level of multicellular organisms has occurred in "hard" artificial life concerned with various forms of autonomous physical agents or robots. This is artificial life's most direct overlap with artificial intelligence. Hard artificial life tries to synthesize autonomous adaptive and intelligent behavior in the real world. It contrasts with traditional artificial intelligence and robotics by exploiting biological inspiration whenever possible, and also by aiming to synthesize behaviors characteristic of much simpler organisms than humans. One of the tricks is to let the physical environment be largely responsible for generating the behavior. Rather than relying on an elaborate and detailed internal representation of the external environment, the behavior of biologically-inspired robotics quite directly depends on the system's sensory input from its immediate environment. With the right sensory-motor connections, a system can quickly and intelligently navigate in complex and unpredictable environments. This so-called "behavior-based" robotics has been pioneered by Rodney Brooks [1989; 1990; 1991]. The initial successes involved insect-like robots and it has since been extended to humanoid robots [Adams *et al.*,2000]. Another trick is to let the physical materials out of which the robot is embodied to automatically provide as much functionality as possible [Pfeifer and Scheier, 2001].

Even with behavior-based robots, design of intelligent autonomous agents is difficult because it involves creating the right interconnections among many complex components. The intelligent autonomous agents found in nature are all alive, and their design was achieved spontaneously through an evolutionary process. So artificial life uses evolution to design autonomous agents [Cliff *et al.*, 1993]. To this end, genetic algorithms have been used to design many aspects of robots, including control systems and sensors [Nolfi and Floreano, 2000; 2002].

In natural autonomous agents, the control system is tightly coupled with morphology. Sims [1994] showed ten years ago how to recreate this interconnection when he simultaneously coevolved simulated creatures' controllers, sensors, and morphology, but he relied on special-purpose software running on extremely expensive supercomputers. More recent advances in hardware and software have enabled this line of research to be pursued with off-the-shelf software running on laptops [Taylor and Massey, 2001]. This work, like Sims's, involves simulations alone. Jordan Pollack and his students have taken the next step and used similar methods to develop actual physical robots. They have connected simulated co-evolution of controllers and morphology with off-the-shelf rapid prototyping technology, allowing their evolutionary design to be automatically implemented in the real world [Lipson and Pollack 2000; Pollack *et al.*, 2001].

Digital evolution

Implementing evolving systems in software is the most practical and constructive way to study many issues about evolving systems, and this "soft" approach has been a dominant trend in artificial life. One of the first significant achievement of spontaneous evolution in a digital medium was Tierra [Ray, 1992], which is simply a population of simple, self-replicating computer programs that exist in computer memory and consume CPU time. A Tierran genotype consists of a string of machine code, and each Tierran "creature" is a instance of some Tierran genotype. A simulation starts when a single self-replicating program, the ancestor, is placed in computer memory and left to replicate. The ancestor and its descendants repeatedly replicate until computer memory is teeming with creatures that all share the same ancestral genotype. Older creatures are continually removed from memory to create space for new descendants. Errors (mutations) sometimes occur, and the population of programs evolves by natural selection. If a mutation allows a program to replicate faster, that genotype tends to spread through the population. Over time, the ecology of Tierran genotypes becomes remarkably diverse. Quickly reproducing parasites that exploit a host's genetic code evolve, and the co-evolution between hoses and parasites spurs the evolution of parasite-resistance and new forms of parasitism. After millions of CPU cycles of this co-evolutionary arms race, Tierra often contains many kinds of creatures exhibiting a variety of competitive and cooperative ecological relationships.

Life has exhibited a remarkable growth in complexity over its evolutionary history. Simple prokaryotic one-celled life led to more complex eukaryotic one-celled life, which led to multicellular life, then to large-bodied vertebrate creatures with complex sensory processing capacities, and ultimately to highly intelligent creatures that use language and develop sophisticated technology — those creatures at the central focus of cognitive science. Although some forms of life remain evolutionary stable for millions of years (e.g, coelacanths and sharks), the apparently open-ended growth in complexity of the most complex organisms is intriguing and enigmatic. Much effort in artificial life is directed toward creating a system that shows how this kind of open-ended evolutionary progress is possible, even in principle. Digital evolution in Tierra does not do this, for significant evolutionary change eventually peters out. Ray has tried to address these limitations by making the Tierra environments much larger and more heterogeneous and by making the ancestral Tierran creatures significantly more complex (in effect, giving them multiple cell types). By allowing Tierran creatures to migrate from machine to machine over the Internet, looking for unused resources and for more favorable local niches, Ray has found signs that they evolve new types of cells [Ray, 2000]. Furthermore, when Tierra is modified so that creatures are rewarded for performing complex arithmetic operations on numbers they find in their local environment, evolution produces the expected increase in genetic complexity [Adami *et al.*, 2000; Lenski *et al.*, 2003]. However, as with the original version of Tierra, these evolutionary progressions eventually stop.

Hillis [1992] demonstrated that co-evolution can spur evolutionary progress, and co-evolutionary arms races might help drive continual evolutionary progression by continually changing the environment for evolution. But the original and most modified versions of Tierra involve some form of co-evolution and yet the environment eventually becomes essentially stable, so there is probably more to the story. Further progress on open-ended evolution would be aided by quantitative comparisons across different artificial and natural evolving systems. Bedau and Packard and their collaborators have taken a step in that direction by defining and studying evolutionary activity statistics. Comparing data from different artificial and natural evolving systems suggests that there are qualitatively different classes of evolutionary dynamics, and no known artificial system generates the kind of evolutionary dynamics exhibited by the biosphere [Bedau *et al.*, 1997; 1998]. We are apparently still missing critical insights about the mechanisms by which evolution continually creates the new kinds of environments that continually elicit new kinds of adaptations.

3 PHILOSOPHICAL IMPLICATIONS OF ARTIFICIAL LIFE

The scientific and engineering of artificial life has rich implications for a number of broad philosophical issues. This section illustrates these implications for a few philosophical issues.

Philosophy and artificial life are natural intellectual partners, for three reasons. By creating wholly new kinds of life-like phenomena, artificial life continually forces us to reexamine and reassess what it is to be alive, adaptive, intelligent, creative, etc. In addition, both philosophy and artificial life seek to understand phenomena at a level of generality that ignores contingencies and reveals essential natures.

Finally, artificial life's computational methodology is a direct and natural extension of philosophy's traditional methodology of *a priori* thought experiment. Aiming to capture the simple essence of vital processes, artificial life abstracts away as many details of living systems as possible. The resulting artificial life models are thought experiments that are explored by actually syntehsizing instances of the models. Like the traditional armchair thought experiments, artificial life simulations attempt to answer "What if X?" questions, but the premises they pose are too complicated to be understood except by synthesizing them. These synthetic methods are often computational (in soft artificial life), but they sometimes involve constructing novel hardware (in hard artificial life) or even construcing novel systems by biochemical means (in wet artificial life). In each case, the motivation is the same: the behavior of the system cannot be determined except through direct experience. These constructive thought experiments bring a new kind of clarity and constructive evidence to philosophy.

Emergence

One of life's amazing features is how the whole is more than the sum of the parts. This is called emergence [Bedau and Humphries, 2007]. As a general definition, emergent phenomena are macro and micro phenomena that are related so that the macro both depends on and is autonomous from the underlying micro phenomena.

Although apparent emergent phenomena are all around us, the two hallmarks of emergence seem inconsistent or philosophically illegitimate. How can something be autonomous from underlying phenomena if it depends on them? This is the traditional philosophical problem of emergence. A solution to this problem would both dissolve the appearance of illegitimate metaphysics and give emergence a constructive role in scientific explanations of emergent macro phenomena like life and mind.

The aggregate global behavior of complex systems studied in artificial life offers a new view of emergence, so-called "weak" emergence [Bedau, 1997; 2003], in contrast to the "strong" emergence that involves in principle irreducibility of macro from micro [Kim, 1999]. On this view, a system's macrostate is emergent just in case it can be derived from the system's boundary conditions and its micro-level dynamical process but only through the process of iterating and aggregating potentially all of the micro-level effects. This new view explains the two hallmarks of emergence. Micro-level phenomena clearly depend on macro-level phenomena; think of how a bottom-up artificial life model works by driving only the local micro processes. At the same time, macro-level phenomena are autonomous because the micro-level interactions in the bottom-up models produce such complex macro-level effects that the only way to recognize or predict them is by observing macro-level behavior. Weak emergence is common in complex systems found in nature, and artificial life's models also exhibit it. The unpredictability and unexplainability of weak emergent phenomena comes from the myriad, non-linear and context-dependent local micro-level interactions that drive the systems. Emergent phenomena can have causal powers on this view, but only by aggregating micro-level causal powers. There is nothing inconsistent or metaphysically illegitimate about underlying processes constituting and generating phenomena in this way by iteration and aggregation. Furthermore, weak emergence is rampant in scientific explanations of exactly the natural phenomena that apparently involve emergence, like life and mind.

This shows how artificial life will play an active role in future philosophical debates about emergence, as well as related notions like explanation, reduction, and hierarchy. Living systems are a paradigm example of emergent phenomena, and artificial life's bottom-up models generate impressive macro-level phenomena wholly out of micro-level interactions. Artificial life expands our sense of what can emerge from what by constructively exploring what is possible.

Evolution

As noted above, the evolution of life has produced a remarkable growth in complexity. Simple prokaryotic one-celled life lead to more complex eukaryotic single-celled life, which then lead to multicellular life, then to large-bodied vertebrate creatures with sophisticated sensory processing capacities, and ultimately to highly intelligent creatures that use language and develop sophisticated technology. This raises a deep question about evolution's creative potential: Does evolution have an inherent tendency to create greater and greater adaptive complexity, or is the increasing complexity of life just a contingent and accidental by-product of evolution? This question has attracted the attention of both philosophers and biologists.

Stephen Jay Gould [1989] devised a clever way to address this issue: the thought experiment of replaying the tape of life. Imagine that the process of evolution were recorded on a tape. The thought experiment is to rewind the evolutionary process backward in time, erasing the tape, and then playing it forward again but allowing it to be shaped by wholly different contingencies. It is not clear what the outcome of the thought experiment is. Gould himself suggests that "any replay of the tape would lead evolution down a pathway radically different from the road actually taken." He concludes that the contingency of evolution destroys any possibility of a necessary growth in adaptive complexity. Daniel Dennett [1995] draws exactly the opposite conclusion. He argues that complex features like sophisticated sensory processing provide such a distinct adaptive advantage that natural selection will almost inevitably discover it in one form or another. Dennett concludes that replaying life's tape will almost inevitably produce highly intelligent creatures that use language and develop sophisticated technology.

I am dubious about both answers, for the same reason. Gould's thought experiment of replaying the tape of life is exactly the right way to investigate the scope of contingency and necessity in evolution. But neither Gould nor Dennett actually carry out the experiment. Instead, they just speculate about what would happen were one to do so. Extensive experience in artificial life has shown time and again that armchair speculations about the outcome of such thought experiments are highly fallible.

We cannot actually replay life's tape, of course, since we cannot roll back time to an earlier biosphere. But we can do the next best thing and synthesize artificial biospheres that are like the real biosphere in relevant respects, and then observe their behavior. The easiest artificial biospheres to construct are simply software systems. The behavior of vast numbers of instances of these software systems can be observed, and very robust generalizations discovered. Obviously, soft artificial life can constructively contribute to this project for it is precisely in the business of creating and studying such systems.

Of course, there is no way to recreate all the conditions of early life on Earth, including the right environment and distribution of species (including the absence of humans). But replaying life's tape does not require returning to life's actual origin. Instead, the subsequent evolution of an entirely different biosphere would

provide even more information about evolution's inherent creative potential, as long as that biosphere's creative evolutionary potential was sufficiently open. So artificial life software systems that are analogous to Earth's early life in relevant respects could serve to replay life's tape.

It is far from trivial to create systems displaying the richness of real life. In fact, no one has yet devised a system that exhibits the continual open-ended evolution that seems to be happening in the biosphere (recall above). Achieving this goal is a key open problem in artificial life, related to its sixth grand challenge. The final evaluation of conjectures like Gould's and Dennett's about evolution's inherent creativity must await artificial life's progress on replaying the tape of life.

Life

Life seems to be one of the most basic categories of actual natural phenomena. Yet it is notoriously difficult to say what life is, exactly. The fact is that today we know of no set of individually necessary and jointly sufficient conditions for life. Nevertheless, there is broad agreement about the distinctive hallmarks that life forms share. These hallmarks include being complex adaptive organization sustained by metabolic processes, produced by natural selection through a process involving random variation and historical contingency, and producing qualitative and unpredictable phenomena involving unique and variable individuals containing unique macromolecules [Mayr, 1982]. The characteristic coexistence of these hallmarks is striking, and it is a reason to suspect there is a unified explanation of life. But appearances might be deceptive. Vital phenomena might have no unified explanation and life might not be a basic category of natural phenomena. Skeptics like Sober [1992] think that the question of the nature of life, in general, has no interesting answer. But one should retreat to skepticism, if at all, only as a last resort. Those searching for extraterrestrial life, those searching for the origin of life on Earth, and those attempting to synthesize life in artificial media typically are betting that there is an interesting explanation of life in general.

Philosophers from Aristotle to Kant have investigated the nature of life, but philosophers today ignore the issue, perhaps because it seems too scientific. At the same time, most biologists also ignore the issue, perhaps because it seems too philosophical. The advent of artificial life is especially revitalized the question today. One can simulate or synthesize living systems only if one has some idea what life is. Artificial life's aim to discern the essence of life encourages liberal experimentation with novel life-like organizations and processes. Thus, artificial life fosters a broad perspective on life. In the final analysis, the question of the nature of life will be settled by whatever perspective provides the best explanation of the hallmarks that living systems exhibit. Better understanding of how to explain these phenomena will also help resolve a cluster of puzzles about life, such as whether life admits of degrees, how the notion of life applies at different levels in the biological hierarchy, and the relationship between the material embodiment of life and the dynamical processes in which those materials participate [Bedau,

1998]. And artificial life provides a constructive setting in which to explore the empirical implications of different conceptions of life.

Motivated partly by experience in artificial life, Bedau [1996; 1998] has recently argued for the admittedly unintuitive view that life in the most fundamental sense is displayed by a system that is continually exhibiting creative evolution. Organisms would then be explained as alive in a derivative sense, by virtue of their connection with and role in an evolving system. One virtue of the conception of life as evolution is that it explains why Mayr's hallmarks of life coexist in nature. We would expect life to involve the operation of natural selection producing complex adaptive organization in historically connected organisms with evolved genetic programs. The random variation and historical contingency in the evolutionary process explains why living phenomena are especially qualitative and unpredictable and involve unique and variable individuals with frozen accidents like chemically unique macromolecules. This view can also explain why metabolism is so important in living systems, for a metabolism is a physically necessary prerequisite in any system that can sustain itself long enough to adapt and evolve. In addition, this view accounts for four of the main puzzles about life [Bedau, 1998].

There are two main objections to this view of life. First, one might think it is entirely contingent that life forms were produced by an evolutionary process. The Biblical story of Adam and Eve shows that is easy to imagine life forms in the absence of any evolutionary process. But it is not clear that this is anything more than a philosophical fantasy, unrelated to what would actually happen anywhere in the real world. A second objection calls attention to the fact that some evolving systems seem devoid of life. Viruses and prions evolve but are dubiously alive, and cultural and technological evolution provides even starker counterexamples. One response to this sort of worry is to bite the bullet and claim that these kinds of evolving systems actually deserve to be considered to be alive, at least in the primary sense. It is important to realize that the project of uncovering the nature of life is not simply to analyze our concept of life. Our concepts are historical accidents that might be unsuited to the underlying categories in nature. It could turn out that the fundamental process that produces the familiar phenomena of life is essentially the same as the process that produces phenomena that we do not today recognize to involve life. If so, then learning this would reveal a new deep truth about life.

Artificial life has called special attention to the question whether purely digital systems existing in computers could ever literally be alive. This question will be easier to answer once there is agreement about the nature of life; but that agreement should not be expected until we have experienced a much broader range of possibilities. So the debate over whether real but artificial life is possible continues. Some people complain that it is a simple category mistake to confuse a computer simulation of life with a real instance of it [Pattee, 1989]. A flight simulation for an airplane, no matter how detailed and realistic, does not really fly. A simulation of a hurricane does not create real rain driven by real gale-force winds. Similarly, a computer simulation of a living system produces merely a symbolic representation

of the living system. The intrinsic ontological status of this symbolic representation is nothing more than certain electronic states inside the computer (e.g., patterns of high and low voltages). This constellation of electronic states is no more alive than is a series of English sentences describing an organism. It seems alive only when it is given an appropriate interpretation.

But this charge of category mistake can be blunted. Artificial life systems are typically not simulations or models of any familiar living system but new digital worlds. Conway's Game of Life, for example, is not a simulation or model of any real biochemical system but a digital universe that exhibits spontaneous macroscopic self-organization. So, when the Game of Life is actually running in a computer, the world contains a new physical instance of self-organization. Processes like self-organization and evolution are multiply realizable and can be embodied in a wide variety of different media, including the physical media of suitably programmed computers. So, to the extent that the essential properties of living systems involve processes like self-organization and evolution, suitably programmed computers will actually be novel realizations of life.

Mind

All forms of life have mental capacities, broadly speaking [Dennett, 1997]. They are sensitive to the environment in various ways, and this environmental sensitivity affects their behavior in various ways. Furthermore, the sophistication of these mental capacities seems to correspond to the complexity of those forms of life. So it is natural to ask if there is an interesting connection between life and mind. For example, life and mind would be strikingly unified if Beer [1990, p. 11] is right that "it is adaptive behavior, the ... ability to cope with the complex, dynamic, unpredictable world in which we live, that is, in fact, fundamental [to intelligence itself]" (see also [Maturana and Varela, 1987/1992]). Since all forms of life must cope in one way or another with a complex, dynamic, and unpredictable world, perhaps this adaptive flexibility intrinsically connects life and mind. Understanding the ways in which life and mind are connected is one of the basic puzzles about life.

Many mental capacities are certainly adaptations produced by the process of evolution of living organisms. This is sufficient for a certain shallow connection between life and mind. Aristotle's view that there is an intrinsic conceptual unity of life and mind goes much deeper. For Aristotle, an organism's mental activity consists of the exercise of various internal capacities and potentialities (its "soul"), and being alive consists of the exercise of those very same capacities and potentialities [Code and Moravcsik, 1992]. The theory of life as continual creative evolution (recall above) implies a related view, according to which the mind as an expression of a process (creative evolution) that is also the definitive feature of life. One specific way to make this argument is by appealing to the suppleness of life and mind [Bedau, 1977a; 1999].

It is well known in the philosophy of mind and artificial intelligence that the

emergent dynamical patterns among human mental states are especially difficult to describe and explain. Descriptions of these patterns must be qualified by "ceteris paribus" clauses, as the following example illustrates: If someone wants a goal and believes that performing a certain action is a means to that goal, then *ceteris paribus* they will perform that action. For example, if someone wants a beer and believes that there is one in the kitchen, then he will go get one — unless, as the "ceteris paribus" clause signals, he does not want to miss any of the conversation, or he does not want to offend his guest by leaving in midsentence, or he does not want to drink beer in front of his mother-in-law, or he thinks he had better flee the house since it is on fire, etc.

This pattern exhibits a special property that I will call "suppleness". Suppleness is involved in a distinctive kind of exceptions to the patterns in our mental lives — specifically, those exceptions that reflect our *ability to act appropriately* in the face of an open-ended range of contextual contingencies. These exceptions to the norm occur when we make *appropriate* adjustment to contingencies. The ability to adjust our behavior appropriately in context is a central component of the capacity for intelligent behavior.

A promising strategy for explaining mental suppleness is to follow the lead of artificial life, because there is a similar suppleness in vital processes such as metabolism, adaptation, and even flocking. For example, a flock maintains its cohesion not always but only for the most part, only *ceteris paribus*, for the cohesion can be broken when the flock flies into an obstacle (like a tree). In such a context, the best way to "preserve" the flock might be for the flock to divide into subflocks. Artificial life models of flocking exhibit just this sort of supple flocking behavior.

Or consider another example concerning the process of adaptation itself. Successful adaptation depends on the ability to explore an appropriate number of viable evolutionary alternatives; too many or too few can make adaptation difficult or even impossible. In other words, success requires striking a balance between the competing demands for "creativity" (trying new alternatives) and "memory" (retaining what has proved successful). Furthermore, as the context for evolution changes, the appropriate balance between creativity and memory can shift in a way that resists precise and exceptionless formulation. Nevertheless, artificial life models can show a supple flexibility in how they balance creativity and novelty [Bedau, 1999]. The suppleness of both life and mind suggests that they might be two different manifestations of essentially the same kind of underlying process, two sides of the same coin. This suggestion is a very open question today, but it shows how artificial life might deeply unify life and mind.

Ethics

Both the process of pursuing artificial life research and the scientific and practical products of that research process raise complicated ethical issues [Bedau *et al.*, 2000]. These issues include four broad categories: (i) the sanctity of the biosphere, (ii) the sanctity of human life, (iii) the responsible treatment of newly generated

life forms, and (iv) the risks of using artificial life technology.

Artificial life's ethical issues somewhat resemble those concerning animal experimentation, genetic engineering, and artificial intelligence, and the extensive literature on those topics may guide exploration of the ethical issues in artificial life. On the other hand, creating novel forms of life and interacting with them in novel ways will place us in increasingly uncharted ethical terrain.

Perhaps the most vivid ethical issues arise from wet artificial life efforts aimed ultimately at making new forms of life in the laboratory from scratch [Bedau and Parke, 2007]. These efforts can be expected to generate public concern. Some will object that creating artificial cells is unnatural or fails to give life due respect [Kass, 2002; Cho, 1999], or that it involves playing God [Cho, 1999]. One main driver for these ethical concerns is the fact that creating new forms of life will inevitably involve what I call deciding "in the dark" [Bedau and Triant, 2007]. Decisions "in the dark" are those we have to make even though we are largely ignorant about their possible consequences. New and revolutionary technologies, such as genetic engineering and nanotechnology, are allowing us to change our environment at an accelerating rate. Much of this change is being driven by the private economic interests of large international corporations. But the unprecedented nature of these technological innovations makes their implications for human health and the environment extremely difficult to forecast.

Decision theory [Raiffa, 1968; Resnick, 1987] has a well-developed arsenal for confronting what are known as decisions "under risk" and decisions "under ignorance or uncertainty," but it is unequipped to help with decisions in the dark. Decision theory approaches a decision in a given context by tabulating the different possible actions that could be made in that context, determining the likely consequences of each action, determining the likely social utility of each consequence, and then analyzing this table by calculating such things as each action's expected utility. Decisions "under risk" are those in which the likely consequences of the actions are uncertain and can only be assigned a probability, and decisions "under ignorance or uncertainty" are those in which even the probabilities of the consequences are unknown. In both kinds of decisions, however, the consequences of different courses of action can be tabulated. Decisions "in the dark" are different in just this respect: We are ignorant about even the possible outcomes of our actions, so we cannot even construct a decision table. So contemporary decision theory has no advice to offer about such decisions.

Yet technological innovations are increasingly forcing society to make decisions in the dark. Genetic engineering and nanotechnology are two examples. Recombinant DNA technology and advances in self-assembling molecular systems are now realizing undreamt of new bio- and nanotechnologies, and governments in most developed countries are betting vast economic stakes on the bio-nano future. But at the same time, their risks are also causing growing alarm. Genetically modified foods are now anathema throughout Europe, and Bill Joy created a stir when he described the dangers of combining biotechnology with nanotechnology in such things as artificial cells [Joy, 2000]. Because of the revolutionary novelty of these

technologies, it is impossible for us to know the likely consequences of their development. Yet we nevertheless face choices today about whether and how to develop them, whether and how to regulate them, etc. We have to make these decisions in the dark.

Society today has two main methods for tackling decisions in the dark: risk analysis and the Precautionary Principle. Growing out of decision theory, risk analysis is the primary method by which large organizations and public agencies (e.g., the EPA and the FDA) make decisions with major social and economic implications [Morgan and Henrion, 1990; Wilson and Crouch, 2001, Ropeik and Gray, 2002]. For example, top officials in the U.S. Department of Agriculture cited a Harvard Center for Risk Analysis study to justify FDA inaction about mad cow disease. But it is unclear whether risk analysis can adequately overcome decision theory's shortcomings regarding decisions in the dark.

Much contemporary discussion of genetic engineering and nanotechnology is influenced by the Precautionary Principle, which states that we should ban new technologies that might create significant risks even if we lack clear evidence of such risks [Raffensperger and Tickner, 1999; Morris, 2000]. The Precautionary Principle is designed to apply precisely to situations in which society is in the dark, and it is playing an increasing role in international law, appearing in over a dozen international treaties and agreements (e.g., the Rio Declaration from the 1992 United Nations Conference on Environment and Development). But the Principle is controversial because it seems to ignore the possible benefits of new technologies.

The creation of new forms of life from scratch will create exciting new opportunities. It will also create new responsibilities. The choices society will confront will be especially difficult, because they will require deciding in the dark. Philosophers have a special expertise for helping think through these novel and consequential issues raised by wet artificial life.

4 CONCLUSIONS

This brief survey of the scientific and philosophical implications of contemporary artificial life should allay some pervasive misconceptions. The primary activity in artificial life today is not to produce toy models superficially reminiscent of life. Indeed, software creations comprise only one of its three synthetic methods. Artificial life does aim to create life-like behavior in artificial systems, to be sure, but the point of this is to uncover the essential properties of living systems, wherever they might exist in nature. The potential fruits of such insights are not just theoretical; they also promise to unlock the door to what could be literally called "living technology." Pursing this goal involves interdisciplinary collaboration as well as connection with the traditional sciences such as biology and chemistry. Increasingly empirical and rigorous, artificial life has made incremental advances toward a broad and ambitious agenda. But the extent to which it will achieve this agenda remains an open question.

Artificial life is not just science and engineering. It is also an important new tool for philosophy. In fact, the interests and methods of artificial life and philosophy overlap in a natural way, illustrating how the sciences and the humanities can work together in the pursuit of shared goals. If artificial life is successful in creating wholly new forms of life, it will also have a hand in changing the nature of the world in which we live. In any case, it is clear that artificial life will continue to have a significant and distinctively constructive impact on a wide variety of old and new philosophical questions.

BIBLIOGRAPHY

[Adami *et al.*, 2000] C. Adami, C. Ofria, and T. C. Collier. Evolution of biological complexity. *Proc. Nat. Acad. Sci. USA* 97, 4463-4468, 2000.

[Adams *et al.*, 2000] B. Adams, C. Breazeal, R. A. Brooks, and B. Scassellati. *IEEE Intell. Syst. Applic.* 15, 25-31, 2000.

[Bedau, 1996] M. A. Bedau. The nature of life. In Margaret Boden, ed., *The philosophy of artificial life* (New York: Oxford University Press), pp. 332-357, 1996.

[Bedau, 1997] M. A. Bedau. Weak emergence. *Philosophical Perspectives* 11, 375-399, 1997.

[Bedau, 1997a] M. A. Bedau. Emergent models of supple dynamics in life and mind. *Brain and Cognition* 34: 5-27, 1997.

[Bedau, 1998] M. A. Bedau. Four puzzles about life. *Artificial Life* 4: 125-140, 1998.

[Bedau, 1999] M. A. Bedau. Supple laws in biology and psychology. In V. Hardcastle, ed., *Where biology meets psychology: philosophical essays* (Cambridge: MIT Press) pp. 287-302, 1999.

[Bedau, 2003] M. A. Bedau. Downward causation and autonomy in weak emergence. *Principia Revista Inernacional de Epistemologica* 6, 5-50, 2003.

[Bedau *et al.*, 1997] M. A. Bedau, e. Snyder, C. T. Brown, and N. H. Packard. A comparison of evolutionary activity in artificial evolving systems and the biosphere. In *Proceedings of the Fourth European Conference on Artificial Life* (Husbands, P. and Harvey, I., eds.), pp. 125-134, MIT Press, 1997.

[Bedau *et al.*, 1998] M. A. Bedau, E. Snyder, and N. H. Packard. A classification of long-term evolutionary dynamics. In *Artificial Life VI* (Adami, C. et al., eds.), pp. 228-237, MIT Press, 1998.

[Bedau *et al.*, 2000] M. A. Bedau, J. S. McCaskill, N. H. Packard, S. Raasmussen, C. Adami, D. G. Green, T. Ikegami, K. Kaneko, and T. S. Ray. Open problems in artificial life. *Artificial Life* 6, 363-376, 2000.

[Bedau and Humphries, 2007] M. A. Bedau and P. Humphries, eds. *Emergence: contemporary readings in science and philosophy.* MIT Press, 2007.

[Bedau and Parke, 2007] M. A. Bedau and E. Parke. *Our Future with Protocells: The Social and Ethical Implications of the Creation of Living Technology.* Cambridge: MIT Press, 2007.

[Bedau and Triant, 2007] M. A. Bedau and M. Triant. Social and ethical implications of artificial cells. In M. A. Bedau and E. Parke, eds., *Social and ethical perspectives on protocells.* MIT Press, 2007.

[Beer, 1990] R. D. Beer. *Intelligence as adaptive behavior: an experiment in computational neuroethology* (Boston: Academic Press), 1990.

[Brooks, 2001] R. Brooks. The relationship between matter and life. *Nature* 409, 409-411, 2001.

[Cho *et al.*, 1999] M. K. Cho, D. Magnus, A. L. Caplan D. McGee, and the Ethics of Genomics Group. Ethical considerations in synthesizing a minimal genome. *Science* 286, 2087, 1999.

[Cliff *et al.*, 1993] D. Cliff, I. Harvey, and P. Husbands. Explorations in evolutionary robotics. *Adaptive Behavior* 2, 73-110, 1993.

[Code and Moravcsik, 1992] A. Code and J. Moravcsik. Explaining various forms of living. In Nussbaum and Rorty, *Essays on Aristotle's De Anima* (Oxford: Clarendon Press), pp. 129-145, 1992.

[Dawkins, 1976] R. Dawkins. *The selfish gene.* Oxford: Oxford University Press, 1976.

[Dennett, 1995] D. C. Dennett. *Darwin's dangerous idea: evolution and the meanings of life* (New York: Simon and Schuster), 1995.

[Dennett, 1997] D. C. Dennett. *Kinds of Minds: Towards an Understanding of Consciousness*. New York: Basic Books, 1997.

[Fraser *et al.*, 1995] C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A.R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Firtchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Furhmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J.-F. Tomb, B. A. Dougherty, K. F. Bott, P.-C. Hu, T. S. Lucier. S. N. Peterson, H. O. Smith, C. A. Hutchinson III, and J. C. Venter. The minimal gene component of *Mycoplasma genetalium*. *Science* 270, 397-403, 1995.

[Gillis, 2002] J. Gillis. Scientists planning to make new form of life. *Washington Post*, November 21, A01, 2002.

[Gould, 1989] S. J. Gould. *Wonderful life: the Burgess shale and the nature of history* (New York: Norton, 1989).

[Hillis, 1992] D. Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. In *Artificial Life II* (Langton, C.G. et al., eds.), pp. 313-324, Addison-Wesley, 1992.

[Holland, 1975/1992] J. H. Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, University of Michigan Press; second expanded edition, MIT Press, 1975/1992.

[Holland, 1995] J. H. Holland. *Hidden order: how adaptation builds complexity*, Helix Books, 1995.

[Hutchison *et al.*, 1999] C. A. Hutchison III, S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, and J. C. Venter. Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* 286, 2165-2169, 1999.

[Joy, 2000] Joy, B. 2000. Why the future does not need us. *Wired* 8 (April). Available on the web at `http://www.wired.com/wired/archive/8.04/joy.html` [last accessed June 2003].

[Kass, 2002] L. R. Kass. . *Life, liberty, and the defense of dignity: the challenge for bioethics*. San Francisco: Encounter Books, 2002.

[Kim, 1999] J. Kim. Making sense of emergence. *Philosophical Studies* 95, 3-36, 1999.

[Langton, 1989] C. Langton, ed. *Artificial life* (Redwood City, CA: Addison-Wesley), 1999.

[Lenski *et al.*, 2003] R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami. The evolutionary origin of complex features. *Nature* 423, 139-144, 2003.

[Lipson and Pollack, 2000] H. Lipson and J. B. Pollack. Automatic design and manufacture of robotic lifeforms. *Nature* 406, 974-978, 2000.

[Maturana and Varela, 1987/1992] H. R. Maturana and F. J. Varela. *The tree of knowledge: the biological roots of human understanding*, 1987. Revised edition (Boston: Shambhala), 1992.

[Mayr, 1982] E. Mayr. *The growth of biological thought* (Cambridge, MA: Harvard University Press), 1982.

[Menger and Angelova, 1998] F. M. Menger and M. I. Angelova. Giant vesicles: imitating cytological processes of cell membranes. *Accounts of Chemical Research* 31, 789-797, 1998.

[Mitchell, 1996] M. Mitchell. *An introduction to genetic algorithms* (Cambridge: MIT Press), 1996.

[Morgan and Henrion, 1990] M. G. Morgan and M. Henrion. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge: Cambridge University Press, 1990.

[Morris, 2000] J. Morris, ed. *Rethinking risk and the Precautionary Principle*. Oxford: Butterworth-Heinemann, 2000.

[Nielsen *et al.*, 1991] P. E. Nielsen, M. Egholm, R. H. Berg, and O. Buchardt. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254, 1497, 1991.

[Nolfi and Floreano, 2000] S. Nolfi and D. Floreano. *Evolutionary robotics: the biology, intelligence, and technology of self-organizing machines* (Cambridge: MIT Press), 2000.

[Nolfi and Floreano, 2002] S. Nolfi and D. Floreano. Synthesis of autonomous robots through evolution. *Trends in Cognitive Science* 6, 31-37, 2002.

[Pattee, 1989] H. H. Pattee. Simulations, realization, and theories of Life. In C. G. Langton, ed., *Artificial life* (Redwood City, CA: Addison-Wesley), pp. 63-78, 1989.

[Pfeifer and Scheier, 2001] R. Pfeifer and C. Scheier. *Understanding intelligence*, 2nd printing (paperback edition). Cambridge, Mass.: MIT Press, 2001.

[Pohorille and Deamer, 2002] A. Pohorille and D. Deamer. Artificial cells: prospects for biotechnology. *Trends in Biotechnology* 20, 123-128, 2002.

[Pollack *et al.*, 2001] J. B. Pollack, H. Lipson, G. Hornby, and P. Funes. Three generations of automatically designed robots. *Artificial Life* 7, 215-223. 2001.
[Raiffa, 1968] H. Raiffa. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, Mass: Addison-Wesley, 1968.
[Raffensperger and Tickner, 1999] C. Raffensperger and J. Tickner, eds. *Protecting public health and the environment: implementing the Precautionary Principle*. Washington, D.C.: Island Press, 1999.
[Rasmussen *et al.*, 2003] S. Rasmussen, L. Chen, M. Nilsson, and S. Abe. Bridging nonliving and living matter. *Artificial Life* 9, 269-316, 2003.
[Rasmussen *et al.*, 2004] S. Rasmussen, L. Chen, D. Deamer, D. Krakauer, N. H. Packard, P. F. Stadler, and M. A. Bedau. Transitions from nonliving to living matter. *Science* 303: 963-965, 2004.
[Rasmussen *et al.*, 2007] S. Rasmussen, M. A. Bedau, S., L. Chen, D. Deamer, D. Krakauer, N. H. Packard, and P. F. Stadler. *Protocells: Bridging Nonliving and Living Matter*. MIT Press, 2007.
[Ray, 1992] T. S. Ray. An approach to the synthesis of life. In *Artificial life II* (Langton, C.G. et al., eds.), pp. 371-408, Addison-Wesley, 1992.
[Ray, 2000] T. S. Ray. In *Artificial life VII* (Bedau, M.A. et al., eds.), MIT Press, 2000.
[Resnick, 1987] M. Resnick. *Choices: an introduction to decision theory*. Minneapolis: University of Minnesota Press, 1987.
[Rich and Knight, 1991] E. Rich and K. Knight. *Artificial intelligence*, McGraw-Hill, 1991.
[Ropeik and Gray, 2002] D. Ropeik and G. Gray. *Risk: a practical guide to deciding what's really safe and what's really dangerous in the world around you*. Boston: Houghton Mifflin, 2002.
[Rumelhart and McClelland, 1986] D. E. Rumelhart and J. L. McClelland. *Parallel distributed processing: explorations in the microstructure of cognition*, 2 Vols., MIT Press, 1986.
[Sims, 1994] K. Sims. In *Artificial life IV* (Brooks, R. and Maes, P., eds.), pp. 28-39, MIT Press, 1994.
[Sober, 1992] E. Sober. Learning from functionalism — prospects for strong artificial life. In Langton et al. (1992), *Artificial life II* (Redwood City, CA: Addison Wesley), pp. 749-765, 1992.
[Szostak *et al.*, 20011] J. W. Szostak, D. P. Bartel, and P. L. Luisi. Synthesizing life. *Nature* 409, 387–390, 2001.
[Taylor and Massey, 2001] T. Taylor and C. Massey. Recent developments in the evolution of morphologies and controllers for physically simulated creatures. *Artificial Life* 7, 77-87, 2001.
[von Neumann, 1966] J. Von Neumann. *Theory of self-reproducing automata* (Urbana-Champagne: University of Illinois Press), 1966.
[Walde *et al.*, 1994] P. Walde, R. Wick, M. Fresta, A. Mangone, and P. L. Luisi. Autopoietic self-reproduction of fatty acid vesicles. *J. Am. Chem. Soc.* 116, 11649-116454, 1994.
[Wiener, 1948] N. Wiener. *Cybernetics, or control and communication in the animal and the machine*, Wiley, 1948.
[Wilson and Crouch, 2001] R. Wilson and E. A. C. Crouch. *Risk-benefit analysis*. Cambridge: Harvard University Press, 2001.
[Wolfram, 1994] S. Wolfram. *Cellular automata and complexity*, Addison-Wesley, 1994.

This page intentionally left blank

INDEX

adaptation, 9, 14, 30, 40–42, 53, 65, 116, 166, 180, 181, 183, 186–187, 190, 192, 331, 351, 361, 469–471, 529, 530, 580–581, 586, 592, 598
  constraints on, 122
  and facilitated variation, 194
adaptationism, 116–123, 136, 143–144, 359, 529
  and comparative method, 120
  empirical, 118, 121, 123
  explanatory, 120
  hard, 122
  methodological, 119
  soft, 122
adaptiveness, 181
Agassiz, L., 5, 32, 433
agents, autonomous 590
Agricultural Revolution, 9
alleles, 137–139, 223, 254, 257, 259, 269, 272–273, 285, 288–289, 316, 326, 331, 498–499, 509–511, 514
  pseudo-alleles, 275
Allen, C., 534
allogeny, 430
altruism, 67-69, 147–148, 150, 203, 225–226 (See also reciprocal altruism.)
Amundson, R. A., 119, 121, 185, 186, 530
analogy, 12, 429, 430, 432
analysis of variance (ANOVA), 40, 257
Andreasen, R. O., 471–477
Appiah, K. A., 455–459
archetypes, 433, 445
Ariew, A., 576
Aristotle, 404, 432, 458, 528, 595, 597
arms races, 17
Arthur, W., 174–175
artificial
  cells, 588
  intelligence, 586
  life, 599
  selection, 508
atavisms, 441
autocatalysis, 271, 529
autosomality, 51, 58, 59, 60, 260
Auyang, S., 503
Avery, A., 279
axiomatization, 350, 488–500, 502, 504, 505, 506–507

bacteriophage, 275–279, 280
bands and inter-bands, 262
Barash, D., 228, 240
barnacles, 21
Bateson, P., 580–581
Bateson, W., 253–256, 265, 283, 288, 313, 315–316
  presence-and-absence hypothesis 283
Beadle, G., 250, 274, 286, 295
Beatty, J., 113–114, 495
Beckner, M., 490, 503
Bedau, M. A., 592, 596
Beer, R. D., 597
Bell, G., 183
benefits, accidental, 531
Beth, E., 493
binomial names, 419–420, 423–424
Biogenic law, 180
biogeography, 17, 504
biometry, 315, 316, 319–321

birdsong, 569
Bithorax gene complex, 287
Blackburn, S., 545
Boehm, C., 243
Boolean networks, 586
Boorse, C., 538, 540
Bowler, P., 30, 180
Boyd, R., 187, 458
Bradwardine, T., 487
Braithwaite, R., 488, 541
Brenner, S., 251
Brentano, F., 549
bridge principle, 349
Bridges, C. B., 260
Brooks, R., 590
Brown, W. L., 462, 465
Buchfield, J. D., 30
Buller, D. J., 526, 534
Bullock, S., 392
Bulmer, M., 503
Bunge, M., 378
butterfly, 360, 363

Callaerts, P., 381
Cambridge University, 1, 3, 5–7, 9
Campbell, D., 205, 208, 209, 211–213
canalization, 189, 576
  post-developmental, 580
Carnap, R., 488
Carroll, S., 194, 366
Castle, W., 509
catastrophism, 3–5
Cavalli-Sforza, L. L., 345, 472–477
cell and molecular biology, 503
cell differentiation
  and evolvability, 165
cellular automata, 585
Central Dogma, 189, 251, 290
centromere, 262
Chagas disease, 2
Chambers, R., 10, 29
Chetvernikov, S. 54n, 63, 81
chi-squared test, 497
chiasmatype, 266, 269
Chomsky, N., 567
Christianity, 5, 30, 31
chromatids, 270
chromosomes, 259, 509
  abberation of, 262
  closed circle c, 269
  inversions in, 262
  non-disjunction of, 261
  polytenic, 262
  polymorphisms of (Dobzhansky), 289
  translocations of, 262
  Y-, 259
Churchland, P., 555
cistron, 277
cladistics and race, 471–475, 477
Clark, A., 373, 376
classification, 404
  systematic, 456
climate, grand theory of, 4
code
  comma-less, 280
  redundant, 280
  triplet, 280
codon, 280
Collins, J., 577
complementation test, 272
complexity, 68, 97, 104, 118, 120, 148, 164, 176, 240, 281, 393, 591, 594
computer simulations, 585
concepts
  innate, 559
  novel, 559
connectionism, 587
consciousness, 564
consilience, 11, 16, 30, 504
constraint, 185
content, intentional, 549
convergence, 236, 432, 437–438, 441
copy-choice, 268
coral reefs, 4

Correns, C., 253, 310
correspondence rules see semantics
Corsi, P., 29
coupling, 265
Cowie, F., 576
Crick, F. H. C., 250, 270, 276, 278, 290, 349
crystals, aperiodic 273
cross-over, 267
Cummins, R., 531, 536, 557
Cuvier, G., 3
cybernetics, 542
cytogenetics, 261

Darwin, C., 1, 2, 4–6, 10, 13–15, 18, 20–22, 24, 27, 28, 32, 39, 111, 112, 181, 203, 209, 220, 239, 338, 343, 362, 432–434, 438, 534
  and deism, 6
Darwin, E., 1, 2, 6, 29
Darwin, G., 2
Darwinian nihilism, 239
Darwinian Revolution, 28, 31, 32
Darwinism, 31, 41, 181, 202, 210–212, 215, 309, 311, 314, 317, 319, 321, 323
  Social, 26, 28, 31
Davidson, D., 563
Davies, N., 528
Davies, P. S., 532, 536
Dawkins, R., 24, 118, 121, 183, 184, 224, 225, 528
decision theory, 599
deductive systems, 500
DeLisi, C., 376
Dennett, D., 118, 121, 222, 527, 594, 595
Descent of Man, 209
descent with modification, 19
Desmond, A., 8, 30
development, 179–183, 192, 195, 369–376, 378, 380, 395, 539
  and core processes, 194–195
  and evolvability, 169
developmental
  biology, 179, 503
  canalization, 580
  constraints, 121
  mechanics (Entwicklungmechanik), 252
  modules, 194
  molecular biology, 364
  pathways, 431
  plans, 544
  recombination, 192, 193
  systems theorists, 372
  systems theory, 189, 373
Dewey, J., 202, 212–215
disease, 538, 581
divergence, principle of 14
DNA, 93, 96, 105, 134–139, 156, 170, 184, 250–251, 263, 264, 269–270, 276, 278–280, 289–294, 345, 351, 370, 373, 376–377, 385, 389, 391–395, 409, 421, 512, 530, 589, 599
  hybrid, 270
  recombinant, 599
  selfish, 292
Dobzhansky, T., 71, 72, 461
Dollo, L., 440
dominance, 317, 318, 322, 327, 330
  ratio, 322
double helix replication, 276
Downe, 2
Drake, S., 485
Dretske, F., 549, 562, 571
drift, 38, 46, 63, 117, 124, 330
Drosophila melanogaster, 55, 250, 364
Developmental Systems Theory, 189–191
Dupré, J., 477
ecology, 490, 503, 516
  Chicago school of, 142
ecotype, 469–471
Edinburgh University, 1, 3

Edwards, A. W. F., 345, 466, 474
Elton, C., 62
emergence, 593
empiricist, 11
enhancer-sequences, 293
entropy, 342
Ephestia kühnellia, 286
Ephrussi, B., 250
epigenetic
  mechanisms, 179
  process, 287
  norms, 201
epistemology, 22
  and fallibilism, 214
  descriptive vs. prescriptive, 208–209, 213
equality, 230
Ereshefsky, M., 456
Escherichia coli, 251
essentialism, 405–407, 456, 458–460, 466, 475
ethical norms, 201
ethics, 22, 598
euchromatic, 272
Euclidean geometry, 486, 502
eugenics, 43, 46
eukaryotes, 268
evo-devo, 290
evolution, 8, 19, 40, 42–44, 55–71, 112–113, 123, 138, 143, 149, 155, 179–181, 206, 223, 336, 339–345, 472–475, 490, 506, 594
  fragmentation of, 180, 183, 188, 196
  major transitions in, 155
  minimum, 345
Evolution of Epistemic Mechanisms (EEM), 205
Evolution of Epistemic Theses (EET), 205, 206
Evolution of Mechanisms (EM) 206, 207, 212
Evolution of Moral Mechanisms (EMM), 206
Evolution of Moral Theses (EMT), 206
Evolution of Norms (EN), 206, 207, 212
evolutionary
  altruism, 148
  biology, 487
  computation, 587
  epistemology, 203, 206
  ethics, 203, 206, 219, 237
  potential, 173–174
  Synthesis, 141, 142
  Taxonomy, 417
  theory, 37, 206, 490, 495, 502, 503, 506
  transition, 141, 155
evolvability, 174, 191
  as abstraction, 169
exaptation, 117
exons, 293
experimental design, 496, 497
explanation,
  how-possible, 360, 362
  mechanistic, 533
  why-necessary, 360, 362, 367
extended phenotype, 222, 224, 236
extended replicator, 384

Fairfield Osborn, H., 443
fate-mapping, 286
feedback loops, 542
female choice, 21
Fermat, P. de, 335
Feyerabend, P., 350
final cause, 9
finches, 7
first-order predicate logic, 485–489, 490–493, 499, 505
Fisher, R. A., 30, 38–47, 53, 69, 78, 252, 310, 317, 319–332, 336–342, 487, 490, 504, 510: *see*

also Fundamental Theorem of Natural Selection
Fisher–Wright controversy, 45, 46, 53, 77, 340
Fundamental Theorem of Natural Selection, 84, 341
and information, 342
method of maximum likelihood, 336
fitness, 41–43, 112, 146, 184, 330, 336–341, 511, 515, 516, 573
propensity interpretation, 113, 168
Fitzroy, R., 7
Fodor, J., 567
Forber, P., 115
formalisation, 485–488, 490, 491, 499–502, 503, 505, 506, 508, 516
Galilean conception of, 487, 490, 491, 500–506, 519
fossil record, 17
Fox Keller, E., 192
Fraassen, B. van, 487, 493, 495, 501
function, 351, 525, 549, 554
agent model of, 526
analysis of, 531
artefact model of, 526
Causal Role (CR) account of, 531, 554, 556
etiological theory of, 534, 550
and explanation, 557
and goal-directedness, 526, 539
heavy vs. light f. talk, 525
indeterminacy of, 561
and malfunction, 531, 536, 551, 557
Selected Effect (SE) account of, 530, 552, 556
functionalism, 554
in biology, 351
Fundamental Theorem of Natural Selection, 41, 42, 44, 80, 327–329, 336, 339, 340, 341, 342
Gödel, K., 486, 489, 501
Galapagos, 7, 9, 14, 17
Galileo, G., 487, 500
Galton, F., 20, 310–315, 338
Game of Life, 597
game theory, 502
gametes, 508
Gannett, L., 463, 474–476
Garfinkle, A., 354, 356, 357
Gegenbaur, K., 433, 435
Gehring, W. J., 381
gemmules, 19, 20
gene, 183, 250, 351
colinearity of sequence, 277, 296
conversion, 269
differential concept of, 258
eyespot, 366
gene frequencies, 328, 330, 331
gene/environment dichotomy, 570
hedgehog, 366
as 'markers', 258
one g. — one enzyme, 274
recessive, 60
genetic(s), 257, 495
accommodation, 192, 193
agency, 142
algorithm, 586, 590
analysis, 249
canalization, 578
drift, 39, 45, 67
engineering, 290, 599
interactive g. systems, 330, 331
linkage, 264
mechanisms, 179
regulation, 291
variance, 331
genetically modified foods, 599
genic fundamentalism, 151
genocentricity, 257
genotype, 193, 256, 513, 515
frequencies, 499
space, 185
environmental interactions, 284
phenotype map, 185

geology, 3, 4
Gerhard, J. C., 191, 194
germ plasm, 182
Ghiselin, M., 460
Gibson, G., 192
Giere, R., 488
Gilbert, S., 188, 566
Ginsborg, H., 538
Glasgow, J., 455, 467, 476–477
Glen Roy, 4–5
Godfrey-Smith, P., 120, 123, 390–393, 395, 396, 532, 533
Goethe, J. W. von, 29
Golden rule, 24
Goldschmidt, R., 266
Goodman, N., 445
goodness of fit, 496, 497, 512, 513
Goodwin, B., 186, 192
Goudge, T., 490
Gould, J., 7
Gould, S. J., 7, 117, 119, 180, 186, 187, 361, 455–457, 463, 467, 594, 595
Grant, R., 3, 8, 29
Gray, R. D., 189, 190, 373
Greenspan, R. J., 194
Grice, H. P., 550
Griffiths, P. E., 24, 189, 190, 373, 379, 381, 390, 395, 533, 535, 543, 567
growth-rate theorem, 346
gynandromorphic flies, 286

Hadorn, E., 286
Haeckel, E., 434
Haeckel, G. von, 180
Hagedoorn effect, 39
Haldane, J. B. S., 37, 41, 44, 45, 47, 58–86, 487, 504, 510
Halder, G. P., 381
Hamburger, V., 179, 184
Hardy, G. H., 509
Hardy–Weinberg law, 50–52, 54, 56, 80, 123, 274, 288, 317–319, 490, 499, 509
Hartl, D., 504
Harvey, W., 535
health, 538
Hempel, C., 488
Hennig, W., 472
Herbert, S., 3
heredity, 310, 311, 315, 317, 321, 326, 508, 510
  chromosomal theory of, 250, 258
  criss-cross pattern of, 260
Herschel, J. F. W., 6, 11, 13
heterochromatin, 272
heteroduplexes, 270
heterozygote superiority, 53, 59, 118
hierarchy, biological, 429, 431
Hilbert, D., 502
Hillis, D., 592
HMS Beagle, 1, 2, 4, 5, 7
holism, 486
Holland, J., 586
homeobox, 290
homeomorphic, 495, 496, 502
homeotic mutants, 290
homogeny, 436
homology, 504, 536
  deep, 431
  definition, 429
  latent, 430, 443
  and similarity, 430
homoplasy, 429–432, 435, 437, 441, 443
  and parallelism, 430, 438, 441
honey bee, 16
Hooker, J., 2
Hrdy, S. B., 232
Hull, D., 350, 456–458, 460, 462, 465
Human Genome Project, 251
Hume, D., 23
Huxley, A., 2, 22
Huxley, J., 204, 208
Huxley, T. H., 2, 10, 20, 31, 203, 225
hybridization, 66, 68

in situ, 292
in vitro, 289

idealization, 558
independent assortment, 508
individual, 324, 406–410
industrial melanism, 289
Industrial Revolution, 3, 9
information, 277, 342, 551
Fisher vs. Shannon, 342
genetic, 369, 375, 376, 378
intentional, 380, 381
semantic, *see* iformation, intentional390
theory, 378, 379
inheritance, 180, 181, 183, 195, 197, 543
blending, 314, 318
conception-donation model of, 187
two-locus model of, 54
Mendelian, 53, 182
units of, 191
innateness, 189, 567, 571, 579
intentional
information, 390
objects, 560
sciences, 553
intentionality, 549, 564
and error, 552
interactor, 144
intercellulare Pangenesis, 252
interpretation, 486, 491, 492
introls, 293
Irish, 26, 27
isolating mechanism, 71
Is-to-Ought-Principle, 238
isomorphic, 495, 496, 497, 499, 502
Jablonka, E., 187
Jacob, F., 291
Jennings, H. S., 51-52
Jimson weed (Datura), 250
Johannsen, W., 256
Joy, B., 599
just-so stories, 118, 361

Kant, I., 23, 24, 29, 220, 528, 533, 544, 595
Kantian projectivism, 544
Kaplan, J., 458, 463–465, 474
Kauffman, S., 529
Keith, Arthur 55
Kihara, H., 102
Kimura, M., 101
kinds
biological, 455
homeostatic property cluster, 408–410
natural, 404–410, 455–460, 477
kinetic theory, 321, 323, 326, 327
Kirschner, M., 191, 194
Kitano, H., 192, 194
Kitcher, P. K., 353–354, 356, 384, 463–465, 468, 472–474, 476
Knight, R. D., 373
Komai, T., 102
Krebs, J., 538
Kropotkin, P., 203
Kuhn, T., 14, 31, 32, 350, 491
Kumazawa, M., 102

Lamarck, J. B., 6, 7, 8, 14, 29
Lamarck, P. A., 180
Lamarckism, 8
Lamb, M., 187
landscape,
adaptive, 45, 94–95, 176, 329, 330, 340, 515, 516
epigenetic, 575–582
Langton, C., 585
language acquisition, 568
Lankester, E. R., 435, 432
laws, 43, 489, 491, 495, 499
of ancestral heredity, 311–315, 317
of coexistence, 499
of interaction, 499
of succession, 499

of regression, 312
learning theory, 495
least squares method, 498
Lesliematrix, 519
Lewins, T., 550
Lewis, D., 535
Lewis, E. B., 275
Lewontin, R. C., 117–119, 185–187, 361, 459, 463, 465–467, 473, 569
life, 595
likelihood, 342, 344, 345
lineage, 464, 468, 469, 471–476
linkage 52
linkage disequilibrium, 289
linkage-maps, 266
Linnaean hierarchy, 416–425
Livingstone, F. B., 463
Lloyd, E., 495
Locke, J., 23
locus, 519, 510
logical empiricism, 486, 487, 490, 491, 503
logistic equation, 518
logical positivism, 486
logicism, 501
logistic equation, 517
Lorenz, K., 370
Lush, J. L., 103
Lyell, C., 3, 4, 6, 7, 10, 20
L's paradigm, 4
L's theory of climate, 11

Müller, F., 443
Mackintosh, J., 203
macromutations, 289
mad cow disease, 600
Mahner, M., 378
maize (Zea mays), 250
Malay Archipelago, 10
Malthus, T. R., 19, 32
Malthusian, 10, 13
Malthusian parameter, 328–329
Mameli, M., 187, 374, 383, 385, 571
mammalian middle ear ossicles, 443
marriage, 234, 235
Maruyama, T., 104
materialism, eliminative, 555
mating
assortative 51
parasexual, 269
mathematics, 2, 487, 505, 508
matrix algebra, 490
Matthen, M., 115, 526, 535
Maturana, H., 376
maximisation principle, 341, 335
Maynard Smith, J., 119, 183, 186, 378, 393, 503
Mayr, E., 54, 71, 74-79, 359, 456, 457, 460, 543, 581, 596
McLaughlin, P., 544
McMullin, E., 485
meaning, 551
meaning-structure, 491
Meckel, J. F., 444
meiotic drive, 512
Mendel's Laws, 490
Mendel, G., 15, 182, 249, 310, 508
Mendelian factor, 322–324
genes/genetics, 321, 493, 498
Mendelian heredity, 38, 44, 45
Mendelism, 41, 181, 182, 188, 309–311, 314–317, 319, 323
messenger-RNA, 292
metabolism, inborn errors of, 283
method of least squares, 336
method of maximum likelihood, 343, 345
micelles, 590
microsatellites, 263
Midgley, M., 222, 239
migration, 61, 330
Millikan, R., 530, 549, 562
mind, 597
Misak, C., 488
misrepresentation, 550, 552
mockingbirds, 7

model, 323, 324, 326, 328, 332, 487, 490–492, 495, 499, 501–504, 506, 509, 510, 517–519
model pluralism, 152
model theoretic, 486
modern synthesis, 44, 54, 66, 179, 196

modularity, developmental, 172–173
molecular-signposts, 263
Monod, J., 221, 291
monogamous, 230, 233
Moore, G. E., 204
moral
  individualism, 211
  progress, 243
  realism, 211
  theory, 219
Morgan, T. H., 250
Muller, H.J., 250
multicellularity, 163
multiple regression, 313, 314, 317
Munro, E., 191, 194
Munz, P., 209, 213
mutagenesis, chemical, 279
mutation, 60–61, 253, 326, 327, 329
  and colinearity, 277
mutation rate, 166
muton, 277
Mycoplasma genitalium, 589
myxoma virus, 154

Nagel, E., 359, 488
nanotechnology, 599
narrow content, 564
natural selection, 8, 13, 16, 40, 42, 111, 114–116, 123–125, 141, 190, 195, 309, 311, 315, 338, 351, 507, 508
  Darwinian, 38, 44, 45, 251
  models of, 57–58
  theory of, 181, 490, 503–505
natural state model, 459–460
natural theology, 7, 528, 539
Naturalism, 545
naturalistic, 23
naturalistic fallacy, 202, 204, 210
nature versus nurture, 257
Neander, K., 114, 530, 535, 562
Nei, M., 472
nematode (Caenorhabditis elegans), 251
Neumann, J. von, 493, 585
neurophysiology, 556
neuroscience, 553, 555
Neurospora crassa, 251
New Synthesis, 289
Newton, Sir I., 2, 8, 11
Newtonian mechanics, 495
niche construction, 385
Nietzsche, F., 221
Nijhout, F., 365
Nissen, L., 540
nomenclature, 419–421
non-linear dynamical systems, 487
normality, 558
Normative Principle, Basic, 241, 243
norm of reaction (NOR), 284
norms, 201
  control, n 235
  enhancement, n 235
  functional, 551, 557, 558
  semantic, 551, 557, 558
Norton, H. T. J., 51, 60n, 80
nucleosomes, 264
nucleotides, 270

O'Malley, C. D., 485
obsolescence, 222, 224
Ockham's razor, 336, 343, 345
Oken, L., 32
ontogeny, 183
  and development, 172
ontology, 503
Open Question Argument, 238
open reading frame (ORF), 277
operationalist fallacy, 113

organisms, 141, 183, 184, 192
The Origin of Species, 338, 343
origin essentialism, 115
orthogenesis, 65
orthologous sequences, 290
Owen, R., 8, 432–434, 445
Oxford School of Ecological Genetics, 46
Oyama, S., 189, 190

Packard, N. H., 592
Paley, 6, 9
Paley, W., 362, 527, 530
Palmer, A. R., 193
Pangenes, 252
pangenesis, 19
Papineau, D., 549
parasexual mating, 269
parasitism, 222
parsimony, 345
  Darwin's Principle, 343–345
partial determination (in evolutionary dynamics), 503
path analysis, 87, 90–91, 325, 330, 332
path coefficients, 325
peak shifts, 45
Pearl–Verhulst, 517
Pearson, K., 51, 255, 312, 313, 315, 316, 319, 320, 322, 323, 338, 497
penetrance, 264, 285
phase space, 493, 498, 499, 516
pheneticism, 462
phenotype, 193, 256
  p. space, 185
  accommodation, 192
  novelties, 193
  plasticity, 192
physicalism, 348, 351, 352, 356, 357, 359, 554,
physiology, 76, 118, 192, 223, 469, 532, 556, 557
Pietrosky, P., 562
Pigliucci, M., 458, 463–471, 474
Pinker, S., 222, 228, 241
plasticity, 167, 192, 194, 195, 196, 582
  phenotypic, 167, 169, 172, 192, 195, 514
  developmental, 193, 195
  adaptive, 193
Plato, 22, 23, 242, 404, 570
pleiotropy, 66, 119, 121, 173, 264, 286, 535, 537
pluralism,
  and units of selection, 141, 150,
  taxonomic, 411–416
  agent and unit, 152
Pollack, J., 590
polyclones, 287
polygamy, 230, 233
polymorphisms,
  electrophoretic, 289
  Restriction-Fragments-Length-P (RFLP), 263
  single nucleotide p. (SNP), 263
Popper, K., 23, 206
Popperian, 32
population, 181, 317, 318, 320, 321, 323, 325–327, 329, 331
  and evolvability, 170–171
  structure theory, 410–411
  p.-level process, 196
population genetics, 38, 41, 45, 487, 490, 503, 508–510
  mathematical, 37
  Mayr's critique, 74–76, 80–82
  Waddington's critique, 73, 76, 80–82
position effect, 272, 285
potential energy, 365
poverty of stimulus, 573
Precautionary Principle, 600
Prinz, J., 576
prion, 596
process

Mendelian, 182
population level, 181
Prokaryotes, 263
promoter, 293
protein synthesis, 386–389, 392–394
Provine, W., 54, 71, 76, 79, 504
Pairing-and-Separation process, 354, 356, 357
punctuated equilibrium, 289
Punnett, R. C., 51, 80, 317, 319, 323, 509
Putnam, H., 354, 356, 357, 488

quantum mechanics, 495
quasi-linkage equilibrium, 104
Quine, W. V. O., 213, 214, 491

race,
biological concept of, 455–477
biological reality of, 455, 457, 462, 464, 467, 468, 471, 477
cladistic concept of, 471–475, 477
ecological concept of, 464, 468–471
geographical concept of, 460–465, 468, 471
phylogenetic conceptions of, 464, 468–469, 471–477
and typology, 456–461
typological concept of, 456–458, 460
Rachels, J., 202, 210, 214
racialism/racism, 240, 457
Raff, R., 179, 186
rationalism, 11, 16, 583
Ray, T. S., 591
reading-frame, 277
realism
and classification, 455
critical r., 211
reciprocal altruism, 25
recombination, high-frequency (Hfr), 269
recon, 277
reductionism, 353, 358, 367
and disjunction, 352, 356
historical, 357–363
refraction, Sine Law of, 335
regression, 313, 326
replaying the tape of life, 594
replication, 187
replicator, 143, 183–185, 188, 193
representation, 550, 554, 556
reproductive strategies, 227, 228
repulsion, 265
Rescher, N., 212
reversal, 441
Richards, R, J., 19, 26, 29, 208, 488
Richerson, P., 187
Ridley, M., 176, 230, 462, 465
Risch, N., 468, 472, 474
Risher, R. A., 37
risk analysis, 600
robots, 590
Rosenberg, A., 490, 504
Rosenberg, N., 474–475
Rosenblueth, A., 541
Roychoudhury, A. K., 472
RNA, regulatory 294
rudiment, 440, 441
Rudwick, M. J. S., 5
Ruse, M., 23, 24, 26, 29, 31, 209, 211, 239, 490
Russell, B., 106, 488
Rutherford, S., 170

Samules, R., 570
Sarkar, S., 372, 380
Schaffner,K., 491
Schelling, F., 32
Schiffler, S., 555
Schwann and Schleiden's cellular theory, 251
Schwenk, K., 191

Scots, 27
segregation,
Mendelian, 169, 254, 257, 260–270, 284, 310, 316, 318, 320, 322, 341, 490, 508, 510
social 227, 229
selection, 9, 13, 64, 317, 321, 325–329, 331, 510, 511, 514
cumulative, 171
entwinement and units of, 152
local monism and units of, 152
s.-for vs. s-of, 116
s. for male combat, 21
group, 25, 45, 141–157, 410
individual, 24
intensity of, 56, 62
interdemic, 514, 516
intrademic, 45
kin, 149
multilevel view of, 151
sexual, 21
units of, 45, 141–157
and sex, 57, 59
self-organisation, 529
selfish gene, 24
Sellars, W., 214, 486, 487, 492–496, 499, 505
semantic and correspondence rules, 476, 477, 479, 481, 483–485
sequence
of DNA, 278
paralogous, 290
repetitive, 292
Serres, E. R. A., 444
set theory, 487, 491, 449, 494, 498, 500, 505, 506
and models, 504
and predicates, 488
sex roles, 230
sex-determination, 285
sex-linkage, 260
sexual subordination, 242
Sheppard, P. M., 71
Shifting Balance Theory, 45, 46, 62, 64, 329
Shrewsbury, 1
Sidgwick, H., 204
Siegal, M., 194
Simpson, G. G., 438
Sims, K., 590
Skyrms, B., 503
Smedley, A., 455–457, 464, 467
Smith, A., 14, 337
Smith, H., 589
Snapdragon (Antrirhinum), 250
Sober, E., 114, 123, 187, 354, 457–460, 462, 472, 476, 533, 567, 595
sociobiology, 205, 240
somatoplasm, 182
Sommerhof, C., 539
Sommerhof, G., 541
species, 7, 404–416, 456–463, 466, 472, 474
species selection, 147
Spencer, H., 7, 19, 29, 31, 112, 208, 221
splicing, 293
Stadler, L., 272
state space, 487, 493, 498, 499, 500
statistics, 309, 311, 314–316, 320, 323, 497
Stegmann, U. E., 394
Steigerwald, J., 544
Sterelny, K., 24, 191, 382, 384
Stoll, R., 493
struggle for existence, 9
sub-organismal biology, 183, 192
subspecies, 463, 464, 471
superorganism, 142
supervenience, 352, 356, 357
Suppe, F., 487, 488, 493, 495
Suppes, P., 487, 488, 493, 495, 500, 501, 505
Swampman, 563
synapomorphy, 429
syngeny, 430
syntax, 486

systematics, 404

Takahata, N., 105
Tarski, A., 488, 549, 493
Tatum, E., 251
tautology problem, 112–113
taxonomy, 404
teleosemantics, 535, 549, 550, 553
telomeres, 263
Templeton, A. R., 463, 473–475
Tennyson, A., 29
theoretical terms, 489
theory, 485, 486, 489, 503, 505
  Received View, 493
  scientific, 486, 487, 488, 491–496,499, 500, 502, 503, 519
  semantic approach to, 350
  syntactic approach to, 350, 486–488, 493
theory of forces, 123
theory of natural selection, 41
Thomas, R., 173
Thompson, P., 490, 495, 503
Tierra, 591–592
Tierra del Fuegians, 7, 20
topological, 495, 505
topology, 500
Toulmin, S., 23
traits,
  acquired, 189
  group, 154
type/token distinction, 351
transcription factors, 293
transdetermination, 287
Tree of Life, 14, 15
true cause, 11, 16

ultimate/proximate causes, 591
uniformitarian, 3, 4
unit-characters, 249
unitarianism, 6
units of selection. See entries under 'selection'.
Universal Grammar, 558
utilitarianism, 24

Varela, F. J., 376
variance, 40, 42, 183, 321–322, 324, 325, 329, 338–339, 341
  and environmental uniformity, 170
  generation of, 166
  non-additive, 45
  and selection, 166
  structured, 175
variation, 13, 65, 310–313
vehicles, 143, 184
Venter, J. C., 589
verae causae, 11, 13
vesicles, 589
vestige, 440, 441, 504
virus, 596
von Baer's laws, 444
von Baer, K. E., 433, 434, 445
von Bertalanffy, L., 541
von Dassow, G., 191, 194
Vorzimmer, P. J., 15, 30
Vrba, E., 117
Vries, H. de, 252, 310, 315

Waddington, C. H., 73–81, 186, 189, 193, 221, 489, 529, 576
Wagner, G., 173, 186, 191
Wallace, A. R., 10, 12, 21, 22, 25, 29, 112
Walsh, D. M., 186
Watson, J. D., 250, 349
Wedgwood, E., 1, 3, 6, 9
Wedgwood, J., 1, 14
Weinberg, W., 509
Weiner, N., 585
Weisheipl, J. A., 485
Weissman, A., 182, 188, 259
Weissmanism, 181, 188
Weldon, W. F. R., 313, 315
well-formed formula, 488, 489
Wendler, D., 577

West-Eberhard, M. J., 192, 193, 195
Westermarck, E., 209
Westminster Abbey, 2
Whewell, W., 3, 10, 11, 29
Whitehead, A. N., 488
wild type, 260
Williams, G. C., 72, 529, 530
Williams, M., 490, 504, 505
Wilson, E. B., 258
Wilson, E. O., 204, 209, 229, 239, 462, 465
Wimsatt, W., 172
Wittgenstein, L., 202, 203, 215
Wolpert, L., 364
Woodger, J. H., 489
Wright, L., 530
Wright, S., 30, 37–39, 41, 44–47, 50, 51, 54, 67, 70, 75–76, 79, 102, 317, 325–332, 340, 497, 514, 520. (See also Shifting Balance Theory; Landscape, adaptive.)
  and potential function, 340, 341
Wright–Fisher controversy, See Fisher-Wright controversy

X^2, 513
X^2-test, 512
X-chromosomes, 259
X-rays, 262

yeast (Saccharomyces cerevisiae), 251
Yukawa, H., 102
Yule, G. U., 50, 51, 53, 255, 288, 317, 322, 509

zero-force law, 123
Zoonomia, 6, 8, 29